

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:33:05 ; Search time 6805.28 Seconds
(without alignments)
9873.061 Million cell updates/sec

Title: US-10-789-164-1

Perfect score: 1182

Sequence: 1 atggcgctgcttactcattat.....aggtaagcagctcgtaagc 1182

Scoring table: IDENTITY NUC
Gapop 10_0 , Gapext 1.0

Searched: 5883141 segs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seg length: 0

Maximum DB seg length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pac:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_str:*
11: gb_sy:*
12: gb_un:*
13: gb_vl:*
14: gb_hlg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1171.2	99.1	1182	6	ES0425
2	1171.2	99.1	2775	1	DQ086153
3	1171.2	99.1	3045	6	ES0424
4	1169.6	99.0	4315	1	AB039950
5	1166.4	98.7	2989	1	AB016245
6	1166.4	98.7	200690	1	BSUB0019
7	1164.8	98.5	3738	1	AB046355
8	1056.4	89.4	8189	1	BSZ92954
9	772.8	65.4	110000	1	AE017331_36
10	772.8	65.4	110000	1	CP000002_36
11	733.2	62.0	8164	1	BSZ92953
12	548.4	46.4	819	6	AX434279
13	507.2	42.9	3244	1	BACCAPAC
14	507.2	42.9	94829	1	AE011191
15	507.2	42.9	94830	1	AE017335
16	507.2	42.9	96231	1	AF188935
17	430	36.4	110000	1	BA000028_02
18	397	33.6	1336	1	BF060883

19	338.6	28.6	3240	1	AF269703	AF269703 Staphyloc
20	338.6	28.6	3240	6	AR485657	AR485657 Sequence
21	338.6	28.6	3240	6	AX145021	AX145021 Sequence
22	338.6	28.6	110000	1	CP000029_21	Continuation (22 o
23	338.6	28.6	300275	1	AE016751	AE016751 Staphyloc
24	324.4	27.4	110000	1	AP006716_05	Continuation (6 of
25	323.6	27.4	110000	1	AP008934_03	Continuation (4 of
26	246.2	20.8	861	6	AR484742	AR484742 Sequence
27	246.2	20.8	861	6	AX143599	AX143599 Sequence
28	181	15.3	309	6	AX434752	Continuation (9 of
29	160.2	13.6	110000	1	AJ749949_08	Continuation (16 o
30	134.4	11.4	110000	1	AE009951_15	Continuation (26 o
31	127.4	10.8	110000	1	AE017340_25	Continuation (16 o
32	113.6	9.6	304450	1	BX294134	BX294134 Pirellula
33	88.8	7.5	301508	1	AE017292	AE017292 Leptospir
34	85.6	7.2	10771	1	AE011422	AE011422 Leptospir
35	58.2	4.9	7218	6	166494	166494 Sequence 14
36	52.6	4.5	2000	6	AX655393	AX655393 Sequence
37	45.4	3.8	2352	6	AR547539	AR547539 Sequence
38	45.4	3.8	9109	2	CFR09D5	AL023812 Caenorhab
39	44.6	3.8	125020	8	AP429315	AP429315 Homo sapi
40	44.4	3.8	10437	1	AE006060	AE006060 Pasteurel
41	44.2	3.7	199901	14	AC154148	AC154148 Muntiacus
42	43.8	3.7	171218	14	AC156153	AC156153 Bos tauru
43	43.6	3.7	250029	2	AE014820	AE014820 Plasmodiu
44	43.2	3.7	134213	8	CNS05TEV	AL359399 Human chr
45	42.8	3.6	1818	1	AF267210	AF267210 Candidatu

ALIGNMENTS

RESULT 1	ES0425	1182 bp	DNA	linear	PAT 31-JAN-2002
LOCUS	ES0425	Process for producing poly-gamma-glutamic acid.			
DEFINITION	ES0425	1 GI:18629413			
ACCESSION	ES0425	UP 2001017182-A/2.			
VERSION	ES0425	1			
KEYWORDS	UP 2001017182-A/2.				
SOURCE	Bacillus subtilis				
ORGANISM	Bacillus subtilis				
REFERENCE	1 (bases 1 to 1182)				
AUTHORS	Ashinuchi, M., Misono, H. and Soda, K.				
TITLE	Process for producing poly-gamma-glutamic acid				
JOURNAL	Patent: JP 2001017182-A 2 23-JAN-2001;				
COMMENT	NAGASE & CO LTD				
OS	Bacillus subtilis (hay bacillus) IFO 3336				
PN	JP 2001017182-A/2				
PD	23-JAN-2001				
PR	09-JUL-1999 JP 1999196335				
PI	MAKOTO ASHINUCHI, HARUO MISONO, KENJI SODA				
PC	CI2N15/09, COB69/36, CI2N1/15, CI2N1/19, CI2N1/21, CI2N5/10, PC				
CC	CI2N9/00, CI2N9/90				
PH	CI2P13/14, CI2N15/00, CI2N5/00				
FT	Key	Location/Qualifiers			
FT	source	1. 1182			
FT	FT	/organism='Bacillus subtilis (hay bacillus)'			
FT	FT	IFO 3336'			
FEATURES	source	Location/Qualifiers			
ORIGIN	1. 1182				
ORIGIN	/organism='Bacillus subtilis'				
ORIGIN	/mol_type='genomic DNA'				
ORIGIN	/db_xref='taxon:1423'				
ORIGIN	Query Match	99.1%; Score 1171.2; DB 6; Length 1182;			
ORIGIN	Best Local Similarity	99.7%; Pred. No. 5.1e-250;			
ORIGIN	Matches 1173; Conservative	0; Mismatches 3; Indels 0; Gaps 0;			
ORIGIN	7 TGTTACTCATTTATAGCGTGTGCTGTCATCTGTCATCGGATATTAGAAAACGACGA 66				

```
Db      4 TGGTACTCATTTATAGCCTGTGTCTCATACGTGCATCGGAATTTAGAAAAAGACGA 63
Oy      67 CATCAAAAAACATTGATGCCCTCCCTGTTGGGTTGAATTTATACGGCATCCGCGAAAA 126
Db      64 CATCAAAAAACATTGATGCCCTCCCTGTTGGGTTGAATTTATACGGCATCCGCGAAAA 123
Oy      127 TCGACTGTGACAAAGGCTGACAAACCGGAATTTATAGAAAGCCGGTTACAAGCTTTGGA 186
Db      124 TCGACTGTGACAAAGGCTGACAAACCGGAATTTATAGAAAGCCGGTTACAAGCTTTGGA 183
Oy      187 AAAACAAGAGAAACAGATGACAAAGATTTACTGGGACACACCGGAGAAAAAGCCGATT 246
Db      184 AAAACAAGAGAAACAGATGACAAAGATTTACTGGGACACACCGGAGAAAAAGCCGATT 243
Oy      247 AAACGAAACCTCAGGAGCCGCAATATCGAGAGCAAAAAAGAAAGTCATGAGAAACAGTA 306
Db      244 AAACGAAACCTCAGGAGCCGCAATATCGAGAGCAAAAAAGAAAGTCATGAGAAACAGTA 303
Oy      307 GAAAGAGGGGCTAAGCGATTTGTCAATGATGATGCGCTTTAACCCGATTTATCAATC 366
Db      304 GAAAGAGGGGCTAAGCGATTTGTCAATGATGATGCGCTTTAACCCGATTTATCAATC 363
Oy      367 ATCTTTCAAGAAACCTCTGAGAGCCCAATATCGGCGTCATTTGTAATGTTTAGAAGAC 426
Db      364 ATCTTTCAAGAAACCTCTGAGAGCCCAATATCGGCGTCATTTGTAATGTTTAGAAGAC 423
Oy      427 CATATGATGATGATGAGGCGGCGGCTTGAATGAAATTTGCAAGAGCTTTACCGCTCAATT 486
Db      424 CATATGATGATGATGAGGCGGCGGCTTGAATGAAATTTGCAAGAGCTTTACCGCTCAATT 483
Oy      487 CCTTAAATAGGCACTTGTCTCACTTACAGTAACTAGTAATATACCGGATTTCTTTAAACAAA 546
Db      484 CCTTAAATAGGCACTTGTCTCACTTACAGTAACTAGTAATATACCGGATTTCTTTAAACAAA 543
Oy      547 GCAAAAGAAACGAAACAAAGATCATATGCTGATTAATCTCAAAATTTACAGATGATAT 606
Db      544 GCAAAAGAAACGAAACAAAGATCATATGCTGATTAATCTCAAAATTTACAGATGATAT 603
Oy      607 TTACGTAATTTGAATACATGATATTTCCCTGATTAACGCTTCTCTGCGCTGGGTGGCT 666
Db      604 TTACGTAATTTGAATACATGATATTTCCCTGATTAACGCTTCTCTGCGCTGGGTGGCT 663
Oy      667 CAAGGACTCGGATGACGAAAGAAACAGATTTAAGGGAATCTGAATGCGCGCGACAGT 726
Db      664 CAAGGACTCGGATGACGAAAGAAACAGATTTAAGGGAATCTGAATGCGCGCGACAGT 723
Oy      727 CCGGAGACAAATGAAATTTCTTCCGCTGATCAGTCCGAGCGAGCTGGGCACTTTGTTAAT 786
Db      724 CCGGAGACAAATGAAATTTCTTCCGCTGATCAGTCCGAGCGAGCTGGGCACTTTGTTAAT 783
Oy      787 GGGTTTGGCGGAAACGAGCGCTTCTTCTACTTTGAATATATGAAACGTTGTAAGAAATC 846
Db      784 GGGTTTGGCGGAAACGAGCGCTTCTTCTACTTTGAATATATGAAACGTTGTAAGAAATC 843
Oy      847 GGTACCCGAGCCGATGATCCGATCATCATCAAGAACTGCCGGGCGAGACCGGTGCGATCGG 906
Db      844 GGTACCCGAGCCGATGATCCGATCATCATCAAGAACTGCCGGGCGAGACCGGTGCGATCGG 903
Oy      907 ACACAGCAATTCGCAAAATGACGTAATTCCTTATATTTGAAGCAAGTGAATGATCTTAATC 966
Db      904 ACACAGCAATTCGCAAAATGACGTAATTCCTTATATTTGAAGCAAGTGAATGATCTTAATC 963
Oy      967 GGTGAAACAAACGAAACCGATCGTAAAGCCCTATGAAAGGCAAAATTCCTGACAGCAAA 1026
Db      964 GGTGAAACAAACGAAACCGATCGTAAAGCCCTATGAAAGGCAAAATTCCTGACAGCAAA 1023
Oy      1027 CTGCAATGACTTGAATTAAGTCAACAGATGAAATTTATGAAATTTGTTAAAGAAAGATG 1086
Db      1024 CTGCAATGACTTGAATTAAGTCAACAGATGAAATTTATGAAATTTGTTAAAGAAAGATG 1083
Oy      1087 CACAACCGGTGATATATGCGCGCAATTTCAATGCGCGCAGAGCCCTTAATTGAA 1146
```

```
Db      1084 CACAACCGGTGATATATGCGCGCAATTTCAATGCGCGCAGAGCCCTTAATTGAA 1143
Oy      1147 AAATCCAGCAATTAAGGTAAAGCAGCTGTAAGC 1182
Db      1144 AAATCCAGCAATTAAGGTAAAGCAGCTGTAAGC 1179

RESULT 2
DQ086153
LOCUS      2775 bp      DNA      linear      BCT 27-JUN-2005
DEFINITION Bacillus subtilis strain ZU-7 Pgab (pgab), pgac (pgac), and pgaA
            (pgaA) genes, complete cds.
ACCESSION  DQ086153
VERSION    DQ086153
KEYWORDS   DQ086153.1 GI:68138277

SOURCE      Bacillus subtilis
            Bacillus subtilis
            Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE   1 (bases 1 to 2775)
            Shi, F., Xu, Z., and Cen, P.
            Efficient production of poly(glutamic acid) by a new strain
            Bacillus subtilis ZU-7
            Unpublished
JOURNAL     2 (bases 1 to 2775)
            Shi, F., Xu, Z., and Cen, P.
            Direct Submission
AUTHORS     Submitted (04-JUN-2005) Institute of Bioengineering, Department of
            Chemical Engineering and Bioengineering, Zhejiang University, No.38,
            Zheda Road, Hangzhou, Zhejiang 310027, P.R. China
FEATURES
            source
            1..2775
                /organism="Bacillus subtilis"
                /mol_type="genomic DNA"
                /strain="ZU-7"
                /db_xref="taxon:1423"
            1..1182
                /gene="pgab"
            1..1182
                /gene="pgab"
                /function="biosynthesis of poly(glutamic acid)"
            /codon_start=1
            /transl_table=1
            /product="Pgab"
            /protein_id="AAV65940.1"
            /translation="MMLIIITACVILVYIGIEKRRKQKNTDALPVRVNINGIRKSTV
            TRLTGILIEAGYKTVGKTGTDAKMIYWDPEKPIKRFQGFNIGEQKWEETVE
            RGAVALYSECAVNPDIQIIFQBELQANIGATNVVLEDHWDGPTLDELAEATAT
            IPVNGHIVITDSEYTERFKOKAKERNKTVIADNSKITDEYLRFKEVAVFPDNLAL
            GVALGIDESTAFKGMUNAPDPGAMRIPLIISPSEGHVNGFANVADASTINIK
            RYKEIGYTPDPIIIMNCRADRVDRTOGFANDVLPYI EASFLIIGETTERIVKAYE
            GKTPADKLDHLEFYSTBIEIMELKKRMANRVYGVGINHGAEPLEIKIHXYKQQLV
            S"
            1183..1632
                /gene="pgac"
            1183..1632
                /gene="pgac"
                /function="biosynthesis of poly(glutamic acid)"
            /codon_start=1
            /transl_table=1
            /product="Pgac"
            /protein_id="AAV65941.1"
            /db_xref="gi:68138279"
            /translation="MFGSDLIYALILGVLSLIRAEKTIYPAQLVVPYGIYGVNQP
            VFILVLIVLSTIVYIVYKSKPMIYGRKRFAMILITGLVLAIDPLVPIVPEFI
            AEFRIIGIIVPLGIANTIOKQGLTITGSLILSGATFALMFVYLLI"
            1633..2775
                /gene="pgaA"
            1633..2775
                /gene="pgaA"
                /function="biosynthesis of poly(glutamic acid)"
            /codon_start=1
            /transl_table=1
```

ORIGIN

/product="PgaA"
/protein_id="AAV85942.1"
/db_xref="GI:68138280"
/translation="MKKSLSPHEKLLKTKQKKTKNKVFIAPDIVVLFMFAPWAG
KAETPKVTSDDVLSASFVDIMMGRVKEVEKQKDSIFQVYEPFRASDVAGN
FENPVTOKNADKEIHLQTNKESVLEKQKDSIFQVYEPFRASDVAGN
EPKQNDLIVAGKSLDAKKISYOKNGVITATLGTVDYSGKFAKKNTPGLPA
DPEITPMISEKKKADIVVYVSHGQRIENDPNRQKQIARASMDAGADIIIVHHPH
VLEPIEVNGVIFVLSLGFVFDQGMTRDSALVQYHKKNGTRFEVTPIDHHEAT
PAPVKDLSIKQTIIRLEIKDSNFAMKVEDKLPIDIDHSDLKSK"

Query Match 99.1%; Score 1171.2; DB 1; Length 2775;
Best Local Similarity 99.7%; Pred. No. 4.4e-250;
Matches 1173; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 7 TGGTACTCATTTATGACCTGTGCTGTGCATATGTCATGCAATATTAGAAAAAGACGA 66
DB 4 TGGTACTCATTTATGACCTGTGCTGTGCATATGTCATGCAATATTAGAAAAAGACGA 63
QY 67 CATCAGAAAAACATTGATGCTCCCTGCTGTTGGGTGAATATTACGGCATCCGGGAAAA 126
DB 64 CATCAGAAAAACATTGATGCTCCCTGCTGTTGGGTGAATATTACGGCATCCGGGAAAA 123
QY 127 TCGACTGTGACAAAGGCTGACCAACCGGAATATTAGAACCGGTACAAAGCTGTGGA 186
DB 124 TCGACTGTGACAAAGGCTGACCAACCGGAATATTAGAACCGGTACAAAGCTGTGGA 183
QY 187 AAAACAAAGACGAAACAGATCAAGATGATTTACTGGACACACCGGAGGAAAAACCGATT 246
DB 184 AAAACAAAGACGAAACAGATCAAGATGATTTACTGGACACACCGGAGGAAAAACCGATT 243
QY 247 AAAACGAAACCTCAGGGGCGAATATCGAAGCAAAAAAGAACTCATGAGAAACAGTA 306
DB 244 AAAACGAAACCTCAGGGGCGAATATCGAAGCAAAAAAGAACTCATGAGAAACAGTA 303
QY 307 GAAAGAGGGGCTTACCGGATTTGATGATGATGATGATGATGATGATGATGATGATGAT 366
DB 304 GAAAGAGGGGCTTACCGGATTTGATGATGATGATGATGATGATGATGATGATGATGAT 363
QY 367 ATCTTTCAGAGAACTTCGCGAGGCAATATGCGGCTCATGATGATGATGATGATGATGAT 426
DB 364 ATCTTTCAGAGAACTTCGCGAGGCAATATGCGGCTCATGATGATGATGATGATGATGAT 423
QY 427 CATATGATGTCATGAGGGCGGACGCTTGTATGAATTTGACAGACGTTTACCGGTACATT 486
DB 424 CATATGATGTCATGAGGGCGGACGCTTGTATGAATTTGACAGACGTTTACCGGTACATT 483
QY 487 CCTTATATGCGCATCTTGTCTATTAAGATGATGATGATGATGATGATGATGATGATGAT 546
DB 484 CCTTATATGCGCATCTTGTCTATTAAGATGATGATGATGATGATGATGATGATGATGAT 543
QY 547 GCAAAAAGACGAAACCAAAAGTCATGATGATGATGATGATGATGATGATGATGATGATGAT 606
DB 544 GCAAAAAGACGAAACCAAAAGTCATGATGATGATGATGATGATGATGATGATGATGATGAT 603
QY 607 TTAAGTAAATTTTGAATATCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 666
DB 604 TTAAGTAAATTTTGAATATCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 663
QY 667 CAAGCACTCGGATTCGAGAAACGATTTTAAAGGAATGCTGAATGCGCGCGCGAT 723
DB 664 CAAGCACTCGGATTCGAGAAACGATTTTAAAGGAATGCTGAATGCGCGCGCGAT 723
QY 727 CCGGAGCAATGAGAAATCTTCGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 786
DB 724 CCGGAGCAATGAGAAATCTTCGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 783
QY 787 GGGTTTCCGCAACGACGCTTCTTCTACTTGAATATGAGAAACGCTGTAAAGAAATC 846
DB 784 GGGTTTCCGCAACGACGCTTCTTCTACTTGAATATGAGAAACGCTGTAAAGAAATC 843
QY 847 GGTTCACCGACGATGATCCGATCATATGAATGCGCGCGACGACGCTGTGATCGG 906

DB 844 GGTTCACCGACGATGATCCGATCATATGAACTGCGCGACGACCGCTGTGATCGG 903
QY 907 ACACGCAATTCGCAAAAGAGATTTGCTTATTTGAAGCAAGCACTGATCTTAATC 966
DB 904 ACACGCAATTCGCAAAAGAGATTTGCTTATTTGAAGCAAGCACTGATCTTAATC 963
QY 967 GGTGAAACAAAGAACCGATGTAAGAACCTATGAGAAAGCAAAATTCCTGCAGACAA 1026
DB 964 GGTGAAACAAAGAACCGATGTAAGAACCTATGAGAAAGCAAAATTCCTGCAGACAA 1023
QY 1027 CTGCAATGACCTAGATATAGTCAACAGATGAAATTTATGAAATTTTAAAGAAAGATG 1086
DB 1024 CTGCAATGACCTAGATATAGTCAACAGATGAAATTTATGAAATTTTAAAGAAAGATG 1083
QY 1087 CACAAACCTGTCATATATGAGGCTCGGCAATATTCATGCTGCGGACGACCTTTAATGAA 1146
DB 1084 CACAAACCTGTCATATATGAGGCTCGGCAATATTCATGCTGCGGACGACCTTTAATGAA 1143
QY 1147 AAAATCCAGAAATACAGATTAAGGAGCTCTGTAAC 1182
DB 1144 AAAATCCAGAAATACAGATTAAGGAGCTCTGTAAC 1179

RESULT 3
E50424 3045 bp DNA linear PAT 31-JAN-2002
LOCUS E50424
DEFINITION Process for producing poly-gamma-glutamic acid.
ACCESSION E50424
VERSION E50424.1 GI:18629412
KEYWORDS JP 2001017182-A/1.
SOURCE Bacillus subtilis
ORGANISM Bacillus subtilis
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE 1 (bases 1 to 3045)
AUTHORS Ashiuchi, M., Misono, H. and Soda, K.
TITLE Process for producing poly-gamma-glutamic acid
JOURNAL Patent: JP 2001017182-A 1 23-JAN-2001;
NAGASE & CO LTD
COMMENT OS Bacillus subtilis (hay bacillus) IFO 3336
PN JP 2001017182-A/1
PD 23-JAN-2001
PF 09-JUL-1999 JP 1999196335
PR MAKOTO ASHIUCHI, HARUO MISONO, KENJI SODA
PI C12N15/09, C08G69/36, C12N1/15, C12N1/19, C12N1/21, C12N5/10, PC
PC C12N9/00, C12N9/90,
PC C12P13/14, C12N15/00, C12N5/00
CC FT
FH Key Location/Qualifiers
FT source 1..3047 /organism="Bacillus subtilis (hay bacillus)"
FT IFO 3336
FT FT Location/Qualifiers
FT 1..3045 /organism="Bacillus subtilis"
FT /mol_type="genomic DNA"
FT /db_xref="taxon:1423"

ORIGIN

Query Match 99.1%; Score 1171.2; DB 6; Length 3045;
Best Local Similarity 99.7%; Pred. No. 4.4e-250;
Matches 1173; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 7 TGGTACTCATTTATGACCTGTGCTGTGCATATGTCATGCAATATTAGAAAAAGACGA 66
DB 22 TGGTACTCATTTATGACCTGTGCTGTGCATATGTCATGCAATATTAGAAAAAGACGA 81
QY 67 CATCAGAAAAACATTGATGCTCCCTGCTGTTGGGTGAATATTACGGCATCCGGGAAAA 126
DB 82 CATCAGAAAAACATTGATGCTCCCTGCTGTTGGGTGAATATTACGGCATCCGGGAAAA 141
QY 127 TCGACTGTGACAAAGGCTGACCAACCGGAATATTAGAACCGGTACAAAGCTGTGGA 186

```

Db      142  TCGATGTGACAAAGGCTGACAAACGGGAATATTAAATGAAGCCGGTTACAAAGCTTTGGA 201
Qy      187  AAAACAACGAGAACGATGCAAGAAATGATTTACTGGACACACCGGAGGAAAAAGCCGATT 246
Db      202  AAAACAACGAGAACGATGCAAGAAATGATTTACTGGACACACCGGAGGAAAAAGCCGATT 261
Qy      247  AAACGGAAACCTCAGAGGCGCGAATTCGGAGAGCAAAAAGAAAGTATGAGAGAAACAGTA 306
Db      262  AAACGGAAACCTCAGAGGCGCGAATTCGGAGAGCAAAAAGTATGAGAGAAACAGTA 321
Qy      307  GAAAGAGGGGCTTACGCGATTGTCAATGATGATGATGATGATGATGATGATGATGATGAT 366
Db      322  GAAAGAGGGGCTTACGCGATTGTCAATGATGATGATGATGATGATGATGATGATGATGAT 381
Qy      367  ATCTTTCAAGAGAACTTCTGAGAGCCCAATTCGGCGCTATTTGTGAATGTTTAAAGAC 426
Db      382  ATCTTTCAAGAGAACTTCTGAGAGCCCAATTCGGCGCTATTTGTGAATGTTTAAAGAC 441
Qy      427  CATATGATGTATGAGGGGCGAGCGCTTGATGAAATTTGCAAGCGTTTACCGCTCAATT 486
Db      442  CATATGATGTATGAGGGGCGAGCGCTTGATGAAATTTGCAAGCGTTTACCGCTCAATT 501
Qy      487  CCTTATTAATGCGCATCTTGTCAATTAAGATGATGATGATGATGATGATGATGATGATGAT 546
Db      502  CCTTATTAATGCGCATCTTGTCAATTAAGATGATGATGATGATGATGATGATGATGATGAT 561
Qy      547  GCAAAAAGAACGAAACAAACAAAGTCAATCTTGTCTGATTAATCAAAAATTACAGATGAT 606
Db      562  GCAAAAAGAACGAAACAAACAAAGTCAATCTTGTCTGATTAATCAAAAATTACAGATGAT 621
Qy      607  TTACGTAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 666
Db      622  TTACGTAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 681
Qy      667  CAAGCACTCGGCAATTTGACGAAAGAACAGCAATTTAAAGGAATCTGATGATGATGATGAT 726
Db      682  CAAGCACTCGGCAATTTGACGAAAGAACAGCAATTTAAAGGAATCTGATGATGATGATGAT 741
Qy      727  CCGGAGCAATGAGAAATTTCTTCGCTGATCACTCGGAGCGCTGGGCACTTTGTTAAT 786
Db      742  CCGGAGCAATGAGAAATTTCTTCGCTGATCACTCGGAGCGCTGGGCACTTTGTTAAT 801
Qy      787  GGGTTTCCGCAAGCAAGCGCTTCTCTACTTTGAATATATGAAACGCTGTAAGAAATC 846
Db      802  GGGTTTCCGCAAGCAAGCGCTTCTCTACTTTGAATATATGAAACGCTGTAAGAAATC 861
Qy      847  GGTACCCGAGCGATGATCGATCATCATCATGAACTGCGCGCAGACCGGTGTCATCGG 906
Db      862  GGTACCCGAGCGATGATCGATCATCATCATGAACTGCGCGCAGACCGGTGTCATCGG 921
Qy      907  ACAAGCAATTTGGCAATGAGATGCTTATTTATTTGAAGCAAGTGAATGATCTTAATC 966
Db      922  ACAAGCAATTTGGCAATGAGATGCTTATTTATTTGAAGCAAGTGAATGATCTTAATC 981
Qy      967  GGTGAACAACAAGAACCGATGTAAGAGCTATGAAAGAGGCAAAATTCCTGACAGCAAA 1026
Db      982  GGTGAACAACAAGAACCGATGTAAGAGCTATGAAAGAGGCAAAATTCCTGACAGCAAA 1041
Qy      1027  CTGCACTGACCTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1086
Db      1042  CTGCACTGACCTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1101
Qy      1087  CACAACCGGTGTCATATATGAGCGTGGCAATTTATCATGTGCGCGACAGCCCTTTAATTGAA 1146
Db      1102  CACAACCGGTGTCATATATGAGCGTGGCAATTTATCATGTGCGCGACAGCCCTTTAATTGAA 1161
Qy      1147  AAAATTCAGATTAACAAGTAAAGCGCTTCGTAAGC 1182
Db      1162  AAAATTCAGATTAACAAGTAAAGCGCTTCGTAAGC 1197

```

```

AB039950
LOCUS      AB039950      4315 bp      DNA      linear      BCT 12-SEP-2000
DEFINITION Bacillus subtilis capB, capC, capA, ywC genes, complete cde.
ACCESSION AB039950
VERSION   AB039950.1 GI:10119860
KEYWORDS  CAPA; CapC; CapB.
SOURCE    Bacillus subtilis
ORGANISM  Bacillus subtilis
REFERENCE 1 (bases 1 to 4315)
AUTHORS   Tran,L.P. and Itoh,Y.
TITLE      Nucleotide sequence of the capBCA operon in Bacillus subtilis
(natco)
JOURNAL    Published Only in Database (2000)
REFERENCE 2 (bases 1 to 4315)
AUTHORS   Tran,L.P. and Itoh,Y.
TITLE      Direct Submission
JOURNAL    Submitted (08-MAR-2000) Yoshifumi Itoh, National Food Research
Institute, Japan (E-mail:yosifumi@nfri.affrc.go.jp,
Tel:+81-298-38-8075, Fax:+81-298-38-7996)
FEATURES
     source
         1..4315
            /organism="Bacillus subtilis"
            /mol_type="genomic DNA"
            /strain="Aeahikawa"
            /db_xref="taxon:1423"
         941..2122
            /gene="capB"
            /gene="capB"
            /gene="capB"
            /codon_start=1
            /transl_table=1
            /product="capB"
            /protein_id="BAB13484.1"
            /db_xref="GI:10119861"
            /translation="MMLILICAVILVIGILEKRRHOKNIDALPVRVNINGIRGKSTV
            TRLTGILIEAGYKVTGKTGTDBARMIMYDPEEKPIKRPQGNIGQKEMETVE
            RGNANIVSECAVNPVDIOIIFOEELONIGIVNVFEDHVDNGPTLDEIAEATAT
            IPNGHLYITSEYTEPFKQAKERNKRVIIANDSKITDEYLRKEVWPFNDNISAL
            GVAQMLGIDESTAFKGMNAPDPDPAKILILIPSEBGEVYVNGPANDASSTANIYK
            RVKEIGPTDPIIIMNCRADRVPTQFADVLPYIEASLILIGETTEPVIAYEE
            GKIPADKLHDLREKSTDEIMELKKRMHNRVIYGVNIGHAEPLEIKIHBYKXQVL
            S"
         2137..2586
            /gene="capC"
            /gene="capC"
            /gene="capC"
            /codon_start=1
            /transl_table=1
            /product="capC"
            /protein_id="BAB13485.1"
            /db_xref="GI:10119862"
            /translation="MFGSDLYIALILGVLSLIFAEKTVIPAGLVVPYGLVFNOP
            VFILVILVSLTIVIVKYGLSKFMILYGRKRFAMLIITGLVLAIFDLPVIVFEI
            AEFRIIGIIVGLIANTIOKQGLTTFGSTLLLSGATPAIMFYVLI"
         2605..3747
            /gene="capA"
            /gene="capA"
            /gene="capA"
            /codon_start=1
            /transl_table=1
            /product="capA"
            /protein_id="BAB13486.1"
            /db_xref="GI:10119863"
            /translation="MKKELSFHEKLLKLTQOKKTKNGVFIPIVFLMFAEMWAG
            KAEPYKVTSDVDVLSAFVGDIMMGRVVEKVTQKADSIFQYVEPIFASDVYAGN
            FENPVTYOKNROADKEITHLOTNKESVVLKDNMFTVNSANNMADYGVGOMKOTLG
            EPAKNDIVAGSYLSIPAKKISIOYKNGVITIAITGETSVSGSGFAKKNTPGVLP
            DPEITPMISAKKHADIIVVQSHGQETDNDPDROQLRAASDADADIIIVGHHP
            VLEPIEVNGTVIFYSIGNFVFDQGWRTSDSALVQVHLKNGTGRFEVTPIDIHEAT
            PAPVKDSLKOKTIIIRBLTSDNFMKVEDKLTFFDIDHSDLSKX"
         3762..3929

```

gene

CDS

/gene="ymc"
3762..3929
/gene="ymc"
/codon_start=1
/transl_table=11
/protein_id="BAB13487.1"
/db_xref="GI:10119864"
/translation="MKPKAIWPFVAVALVPMFMSAFKFNQDLTDQEKQKIDMENKIK
QOQEEFVANK"

ORIGIN

Query Match 99.0%; Score 1169.6; DB 1; Length 4315;
Best Local Similarity 99.7%; Pred. No. 9.3e-250;
Matches 1172; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

7 TGGTACTCATATTATAGCCCTGCTGTCATATGTCATCGAATAATTAGAAAAAGACGA 66
1064 TGGTACTCATATTATAGCCCTGCTGTCATATGTCATCGAATAATTAGAAAAAGACGA 1003

67 CATCAGAAAAACATTGATCCCTCCCTGTTGGGTTGAATTTAAGCGCATCCGCGAANA 126
1004 CATCAGAAAAACATTGATCCCTCCCTGTTGGGTTGAATTTAAGCGCATCCGCGAANA 1063

127 TCGACTGTGACAAAGCTGACCAACCGGAATTAATAGAGCGGTTACAAAGCTGTTGA 186
1064 TCGACTGTGACAAAGCTGACCAACCGGAATTAATAGAGCGGTTACAAAGCTGTTGA 1123

187 AAAACAACAGAAACAGATGCAAGAAATGATTTACTGGAGACACCGGAGGAAAAAGCGATT 246
1124 AAAACAACAGAAACAGATGCAAGAAATGATTTACTGGAGACACCGGAGGAAAAAGCGATT 1183

247 AAACGGAAACCTCAGGGGCGGAATATCGAGAGCAAAAAAGAGTCAATGAGAGAACAGTA 306
1184 AAACGGAAACCTCAGGGGCGGAATATCGAGAGCAAAAAAGAGTCAATGAGAGAACAGTA 1243

307 GAAAGAGGGGCTTACCGGCTTGTGATGATGATGATGATGATGATGATGATGATGATGAT 366
1244 GAAAGAGGGGCTTACCGGCTTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1303

367 ATCTTCAGAGAAACCTTGTGACAGGCAATATGCGGCTGATTTGTGATGATTTAGAGAC 426
1304 ATCTTCAGAGAAACCTTGTGACAGGCAATATGCGGCTGATTTGTGATGATTTAGAGAC 1363

427 CATATGATGTCATGAGGCGGACGCTTGTGATGATGATGATGATGATGATGATGATGAT 486
1364 CATATGATGTCATGAGGCGGACGCTTGTGATGATGATGATGATGATGATGATGATGAT 1423

487 CCTTATTAATGGCCATCTTGTGATTAACAGATGATGATGATGATGATGATGATGATGAT 546
1424 CCTTATTAATGGCCATCTTGTGATTAACAGATGATGATGATGATGATGATGATGATGAT 1483

547 GCAAAAGAAAGCAACAAAGATCATGATGATGATGATGATGATGATGATGATGATGATGAT 606
1484 GCAAAAGAAAGCAACAAAGATCATGATGATGATGATGATGATGATGATGATGATGATGAT 1543

607 TTAAGTAATTTTGAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 666
1544 TTAAGTAATTTTGAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1603

667 CAAGCACTCGGCAATGACGAAGAAACGCAATTTAAGGGAATGCTGAATGCGCGCCAGAT 726
1604 CAAGCACTCGGCAATGACGAAGAAACGCAATTTAAGGGAATGCTGAATGCGCGCCAGAT 1663

727 CCGGAGCAATGAGAAATCTTCCGCTGATAGTCCGACGAGCGCTGGGCACTTTGTTAAT 786
1664 CCGGAGCAATGAGAAATCTTCCGCTGATAGTCCGACGAGCGCTGGGCACTTTGTTAAT 1723

787 GGGTTGGCCGCAACGACGCTTCTTCTACTTGAATATGAAACGTTGAAAAAATC 846
1724 GGGTTGGCCGCAACGACGCTTCTTCTACTTGAATATGAAACGTTGAAAAAATC 1783

847 GGTATCCGACCGATATCGGATATCATGAACTCCGCGGACGCTGTGATCGG 906

Db 1784 GGTATCCGACCGATATCGGATCATCATCAATGAACTGCCGCGACGCTGTGATCGG 1843

Qy 907 AACAGCAATTTGGCAATAGATGATGCTTATATTTGAAGCAAGTGAATGATCTTAATC 966

Db 1844 AACAGCAATTTGGCAATAGATGATGCTTATATTTGAAGCAAGTGAATGATCTTAATC 1903

Qy 967 GGTGAAACAGCAACCGATGTAAGAACCTTATGAGAGGCAAAATTTCTGCAGACAAA 1026

Db 1904 GGTGAAACAGCAACCGATGTAAGAACCTTATGAGAGGCAAAATTTCTGCAGACAAA 1963

Qy 1027 CTGCATGACCTAGAGTATAGTCAACAGATGAATTAATGAAATTTTAAAGAAATG 1086

Db 1964 CTGCATGACCTAGAGTATAGTCAACAGATGAATTAATGAAATTTTAAAGAAATG 2023

Qy 1087 CACACCGCTGATATATGAGGCTGGCAATTTATGATGCTGCGACGACCTTTAATTGA 1146

Db 2024 CACACCGCTGATATATGAGGCTGGCAATTTATGATGCTGCGACGACCTTTAATTGA 2083

Qy 1147 AAATCCAGAAATACAGTAAGTAAGCAAGCTCTGAAC 1182

Db 2084 AAATCCAGAAATACAGTAAGTAAGCAAGCTCTGAAC 2119

RESULT 5
AB016245 2989 bp DNA linear BCT 14-OCT-1999
LOCUS AB016245 2989 bp DNA linear BCT 14-OCT-1999
DEFINITION Bacillus subtilis pgsb, pgsb and pgsC genes, complete cds.
ACCESSION AB016245
VERSION AB016245.1 GI:6045071
KEYWORDS pgsb, pgsC.
SOURCE Bacillus subtilis
ORGANISM Bacillus subtilis
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE
1 (sites)
Ashiuchi, M., Soda, K. and Misono, H.
A poly-glutamate synthetic system of Bacillus subtilis IFO 3336: gene cloning and biochemical analysis of poly-glutamate produced by Escherichia coli clone cells
Biochem. Biophys. Res. Commun. 263 (1), 6-12 (1999)
10486244
2 (bases 1 to 2989)
Ashiuchi, M.
Direct Submission
Submitted (14-JUL-1998) Makoto Ashiuchi, Kochi University, Research Institute of Molecular Genetics; Otsu 200, Monobe, Nankoku, Kochi 783-8502, Japan (E-mail:ashiuchi@img.kochi-u.ac.jp, Tel:0888-64-5215, Fax:0888-64-5109)
Location/Qualifiers
1.2989
/organism="Bacillus subtilis"
/mol_type="genomic DNA"
/strain="IFO 3336"
/sub_species="Bacillus natto"
/db_xref="taxon:1423"
1.1182
/gene="pgsB"
1.1182
/gene="pgsB"
/codon_start=1
/transl_table=11
/product="pgsB"
/protein_id="BA085263.1"
/translation="MWLLIACAVILVIGLEKRRHKNIDALPVAVNINIRGKSTV
TRLTGILIEAGYKTVGKTGTDAMIWDPFEKPKIKRPGPSIGQKEVARETVE
RGNAIVSCEMVNPYQIIFOBELLQANIVVWLEDMVMGPTLDEIABAFTAT
IYNGHLVITDESEYEFKOKAKERTKVIIDNSKITDEYLRKFEYVFPDNASIAL
GVAQALGIDERTAFKMLNAPPPDGRAMTILPLISPEGFVNGFANASSSTANIK
RVEIGVPTDDPTIINNCRADRVDRTOQANDVLPITASELLIGETTERPIVKAER
GKIPADKADLEKSTDEIMELIKOSMNRVLIYGVNIGAAEPLEIKETIKVKOLV
S"
1197..1646
/gene="pgsC"

```

CDS
1197..1646
    /gene="pgsC"
    /codon_start=1
    /transl_table=1
    /product="PgsC"
    /protein_id="BAA85264.1"
    /db_xref="GI:6045073"
    /transl_table="MFGSDILYALILGVLLSLIPEAKTGIVPAGLVVGEVGLVFNQFQ
VILVILVSLTLYTVKRGSLKFMILVGRKRFAMLLTGVLKIAFPFLVIVPEEL
AEFRGIGIVPCLINLTIQKQSLTTPSTLLSGATFALINFYVLLI"
1665..2807
    /gene="pgsA"
    /gene="pgsA"
    /gene="pgsA"
    /codon_start=1
    /transl_table=1
    /product="PgsA"
    /protein_id="BAA85265.1"
    /db_xref="GI:6045074"
    /transl_table="MKKELSFHEKLLKLTQKQKKKTNGHVFIAIPVFLFAPFMWAG
KAEPTKRVKIVSDVLSASFVGDIMMGRVYEKTEQKADSIFQYVEPIFASDVAGAGV
FENPNLYQKNGKQADKEHILQTNKESVVLKQMFNFTVNSANNHAMDYGVGMQDITGL
EBAKNOIDVGGVAGSLDAKKKISYQXNGVGTIALTGFPIVSGGSGFAKKTPTPLP
DEEITPMISEKKKADIVYVSHNGCEYDNDPNRQQLARMSDAGADIIVGHHPK
VLEIFLVNGVITFIYSLGNFVFDQGMTIRSDALVQHLKNGGTGREFEVTPIDHEAT
PAPVKKDSIQKTIITRELTKOSNFAMKVEDGKLDVQHLIDHSIDLKSK"

```

ORIGIN

Query Match	98.7%;	Score 1166.4;	DB 1;	Length 2989;
Best Local Similarity	99.5%;	Pred. No. 5.1e-249;		
Matches 1170;	Conservative	0;	Mismatches 6;	Indels 0;
			Gaps	0

QY	1	TGGTATCTATTAAATACCTGAGGTGTCATACATACGTTCATCGGAATATTAAAAAAACACCA	86
Db	4	TGGTACTCATTTATAGCCTGTGCTGCATACGTGCATCGGAATATTAAAAAAAGACGA	63
QY	67	CATCGAAAAAACAATGATAGCCCTCCCTGTTCCGGTGAATATTAAACGGCATCCGCCGAAAA	126
Db	64	CATCGAAAAAACAATGATAGCCCTCCCTGTTCCGGTGAATATTAAACGGCATCCGCCGAAAA	123
QY	127	TGCAGCTGTACAAAGCTGACAAACCGGAATATTAAATGAAGCCGGTTAACAGACTGTGGA	186
Db	124	TGCAGCTGTACAAAGCTGACAAACCGGAATATTAAATGAAGCCGGTTAACAGACTGTGGA	183
QY	187	AAAAACAACAGAAACAGATSCAAGATATTACCTGGACACACCGGAGGAAAAACCGATT	246
Db	184	AAAAACAACAGAAACAGATSCAAGATATTACCTGGACACACCGGAGGAAAAACCGATT	243
QY	247	AAACGGAAACCTCAGGGGCCGAATATCGGAGAGCAAAAAAGAGTCATGAGAAACAGTA	306
Db	244	AAACGGAAACCTCAGGGGCCGAATATTGGAAGAGCAAAAAAGAGTCATGAGAAACAGTA	303
QY	307	GAAAGAGGGGCTTAAACGCAATTGTCAGTGAATGCATGGCTTTAACCCAGATTATCAATC	366
Db	304	GAAAGAGGGGCTTAAACGCAATTGTCAGTGAATGCATGGCTTTAACCCAGATTATCAATC	363
QY	367	ATCTTTCAGAGAAACCTTCTGACGGCCAAATATCGGCGTCATTGTGAATGTTTAAAGAC	426
Db	364	ATCTTTCAGAGAAACCTTCTGACGGCCAAATATCGGCGTCATTGTGAATGTTTAAAGAC	423
QY	427	CATATGATGTCAATGGGGCCGACGCTTGAATGAATTCGAAGCGTTTACCGCTACAT	486
Db	424	CATATGATGTCAATGGGGCCGACGCTTGAATGAATTCGAAGCGTTTACCGCTACAT	483
QY	487	CCCTAATATGGCCATCTTGTCATTAACAGATGAGATATACCGAGTTCTTTAAACAAAA	546
Db	484	CCCTAATATGGCCATCTTGTCATTAACAGATGAGATATACCGAGTTCTTTAAACAAAA	543
QY	547	GCAAAAAGAACAAACAAAGTCATCATTTGCTGATACCTCAAAAATTACAGATGAT	606
Db	544	GCAAAAAGAACAAACAAAGTCATCATTTGCTGATACCTCAAAAATTACAGATGAT	603
QY	607	TTACGTAATTTTGAATACATGATATTCCTGATACCGTTCTCTGCGCGTGGGTGTGCT	666

Db	604	TTACGTAAATTGGAAATACATGTAATCCCTGATTAACGCTTCTCGCGCTGGATGGAGCT	663
Qy	667	CAAGCACTCGGCATTTGACGAAGAAACAGCAATTTAAGGAATGCTGAAATGCGCCCGCAGAT	726
Db	664	CAAGCACTCGGCATTTGACGAAGAAACAGCAATTTAAGGAATGCTGAAATGCGCCCGCAGAT	723
Qy	727	CCGGAGCAATGAGAAATTCCTTCGCTCATATGATCCGACGAGCCTGGGCACTTTGTTAAT	786
Db	724	CCGGAGCAATGAGAAATTCCTTCGCTCATATGATCCGACGAGCCTGGGCACTTTGTTAAT	783
Qy	787	GGGTTTCCCGCAACGACGCTTCTTCACTTTGAAATATATGAAACGTTGTAAGCAATC	846
Db	784	GGGTTTCCCGCAACGACGCTTCTTCACTTTGAAATATATGAAACGTTGTAAGCAATC	843
Qy	847	GGTTACCCGACCGATGATCCGATATATATATGAAACGCGCGGAGACCGGTGATCGG	906
Db	844	GGTTACCCGACCGATGATCCGATATATATATGAAACGCGCGGAGACCGGTGATCGG	903
Qy	907	ACAACAGCAATTCGCAATGACGTAATGCTTCTTATTTAAGCAATGAACTGATCTTAATC	966
Db	904	ACAACAGCAATTCGCAATGACGTAATGCTTCTTATTTAAGCAATGAACTGATCTTAATC	963
Qy	967	GGTGAACCAACAGAACCGATGCTTAAGCTATGAAGGCAAAATTCCTGCAGACAA	1028
Db	964	GGTGAACCAACAGAACCGATGCTTAAGGCTATCAACAAAGGCAAAATTCCTGCAGACAA	1023
Qy	1027	CTGCATACCTAGAGTAATAGTCAACAGATGAATTAATGAAATTTGTTAAAGAAAGAAATG	1086
Db	1024	CTGCATATATTAGAGTAATAGTCAACAGATGAATTAATGAAATTTGTTAAAGAAAGTAATG	1083
Qy	1087	CACAAACGGTGCATATATATGCGCTGGCAATATTCATGCTGCGCAGAGCCTTTAATTGAA	1146
Db	1084	CACAAACGGTGCATATATATGCGCTGGCAATATTCATGCTGCGCAGAGCCTTTAATTGAA	1143
Qy	1147	AAATTCACGAATTCAGGTAAAGCACTGCTAAGC	1182
Db	1144	AAATTCACGAATTCAGGTAAAGCACTGCTAAGC	1179

RESULT 6
BSUB0019/c

LOCUS BSUB0019 200690 bp DNA linear BCT 18-APR-2005
DEFINITION Bacillus subtilis complete genome (section 19 of 21) : from 3608981
to 3809670.

ACCESSION	Z99122 AL009126
VERSION	Z99122.2 GI:32468830

KEYWORDS	SOURCE	ORGANISM
Bacillus subtilis subsp. subtilis str.	168	
Bacillus subtilis subsp. subtilis str.	168	

REFERENCE
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
1 (bases 1 to 200630)
Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G.,
Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G.,

Aaevedo, V., Bertero, M. G., Beesieres, P., Bolotin, A., Borchert, S.,
Borries, R., Boutsier, L., Brans, A., Braun, M., Briegleb, S. C.,
Bron, S., Brouillet, S., Brusch, C. V., Caldwell, L. B., Capuano, V.,
Carter, N. M., Choi, S. K., Codani, J. J., Connetton, I. F., Cummings, N. J.,
Daniel, R. A., Denizot, F., Devine, K. M., Dusterhoft, A., Ehrlich, S. D.,
Emmerston, P. T., Entlan, K. D., Errington, J., Fabret, C., Gallazzi, A.,
Folger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galazzi, A.,
Galleron, G., Ghim, S. Y., Glaeser, P., Gottreau, A., Goldsight, E. J.,
Grand, G., Gutschmidt, G., Guy, B. Y., Haga, K., Halecn, J., Harwood, C. R.,
Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M. F.,
Ikeya, M., Jones, L., Jorie, B., Karamata, D., Kaashara, Y.,
Klaer-Blanchard, M., Klein, C., Kobayashi, Y., Koetler, P.,
Koningsstein, G., Krogh, S., Kunano, M., Kurita, K., Kaplans, A.,
Lardinois, S., Lauber, V., Lazarevic, V., Lee, S. M., Levine, A., Liu, H.,
Meunier, S., Manuel, C., Medigue, C., Medina, N., Melillo, R. P.,
Mizuno, M., Moesli, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M.,
Ogawa, K., Ogiwara, A., Oudga, B., Park, S. H., Parro, V., Poll, T. M.,
Portetelle, D., Portoullik, S., Prescott, A. M., Presecan, E., Pujic, P.,
Purnelle, B., Rapoport, G., Ray, M., Reynolds, S., Rieger, M.,
Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadate, Y., Sato, T.,

Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J., Sekowska, A., Seror, S. J., Serro, P., Shin, B. S., Soldo, B., Sotokich, A., Tacconi, E., Takagi, T., Takahashi, H., Takemori, K., Takeuchi, M., Tamakoshi, A., Tanaka, T., Terpstra, P., Tognoni, A., Toato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassarotti, A., Viari, A., Wambut, R., Medler, E., Medler, H., Welter, T., Winters, P., Wipat, A., Yamamoto, H., Yaman, K., Yasunoto, K., Yata, K., Yoshida, K., Yoshikawa, H. F., Zimstein, E., Yoshikawa, H. and Danchin, A.

TITLE
The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*

JOURNAL
Nature 390 (6657), 249-256 (1997)

REFERENCE
2 (bases 1 to 200690)

AUTHORS
Kunet, F., Ogaawara, N., Yoshikawa, H. and Danchin, A.

TITLE
Direct Submission

JOURNAL
Submitted (27-JUN-2003) I. Moszer, A. Danchin, Institut Pasteur, Genetique des Genomes Bacteriens, 28 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail: mosze@pasteur.fr, 75724 adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45 68 89 48

COMMENT
On Jul 7, 2003 this sequence version replaced gi:2636029. This entry contains data from release R16.1 of the Subtilist database. Further data on gene annotation and detailed information about changes from previous releases can be found at <http://genolist.pasteur.fr/Subtilist/>.

FEATURES
source
1. 200690
/organism="Bacillus subtilis subsp. subtilis str. 168"
/mol_type="genomic DNA"
/strain="168"
/db_xref="taxon:224308"
complement (114..2987)
/gene="uvrA"
/locus_tag="BSU35160"
complement (114..2987)
/gene="uvrA"
/locus_tag="BSU35160"
/function="excision of ultraviolet light-induced pyrimidine dimers in DNA"
/codon_start=1
/transl_table=11
/product="excinuclease ABC (subunit A)"
/protein_id="CAB15533.1"
/db_xref="GI:2636042"
/db_xref="GOA:O34863"
/db_xref="InterPro:IPR003439"
/db_xref="InterPro:IPR003593"
/db_xref="InterPro:IPR004602"
/db_xref="Subtilist:BG12697"
/db_xref="UniProt/Swiss-Prot:O34863"
/translation="MAMDRIEVGKARAHNKNIDVTPRDLVVVTGLSGSGKSLAFDTYVAGGRVYVESAVAROPFGMDKPVDAIEGLSPALISIDOKTSPNPSTVGTVEITYVRLIYRVGKPHCPKPGKIGTSTITQMDRIIEYPERKLOVLAIVYGRKGAHVYVLEQIRKQGVYRVIRIDGEMALSDITLKKKHSIEVVIDRIYVKGVAARLSDSLETALRGSRWIVDIYGEELMFSEHACPHGFSIGLEPRLSFSPFACPTCDGLGKLEVDADVIYPMQDLSKENAVAPWTPISQYPPQLLAVCTHYGIDMDVPVDLPHQDLYLGSQGDILYRYENDVGOIRGEIOPGVLNRIERYKGTGSDFIRQMGWYSGKCPKCGKYLKKEALAVYIDRHGKLTLSVADALAFKDLTSEKDMQIANLIREIYVERLSFLDKYGLDYITISRAAGTISGEAQRIRLATOIGRSLGVLVILDEPSIGLHQRNDRLISALKNMDCDNTLIIVEHDDTMAADYLDIDPGAGIHGGOVISAQTEPEVEDPNSLTGSLYSGKFIPLPEPRRPDRYIEIKASNNIKKVNAPFLGTATVGVSGSKTLVNEIILKALAKLAKAKRSGSHKEIKGLDHDKVIDIDAPIGRTPRSNPATYGVDDIDVFAQONEAKVYGGKGRSFPVKGRCSEA CRDGIKTIEMHRLPDVYVCEVCHGRVNRRLVYKGSISDVLDMVEALISFENPKIKRKLQTIYDVGLEYITLQOPTITISGEAQRVYKASELHKRSSTRTLYTIDEPTTGLHVDIDRLVQLRVLDNGDVLVIEHNDITIKIADYIVDLGPEGAGGTTVAGSGPEITVEESYTGRIYLPYERDITKMSILKAKETATS"

gene
/gene="uvrB"
/locus_tag="ASU35170"
complement (2995..4980)
/gene="uvrB"

CDS

/locus_tag="BSU35170"
/function="excision of ultraviolet light-induced pyrimidine dimers in DNA"
/note="alternate gene name: dila, uvrA"
/codon_start=1
/transl_table=11
/product="excinuclease ABC (subunit B)"
/protein_id="CAB15534.1"
/db_xref="GI:2636043"
/db_xref="GOA:P37954"
/db_xref="InterPro:IPR001410"
/db_xref="InterPro:IPR001650"
/db_xref="InterPro:IPR001943"
/db_xref="InterPro:IPR004807"
/db_xref="InterPro:IPR006935"
/db_xref="InterPro:IPR009055"
/db_xref="Subtilist:BG10502"
/db_xref="UniProt/Swiss-Prot:P37954"
/translation="MKDRFELVSKYQPGDQPKAIKLVKGIQEGKKHQTLLGATGTGKTFVSNLIKVNKPTLVIAHKKTLAAGLYSEKFEFPNNAVEVSVSYDYQPEAVY PQTDFTRKDSINDEIDKHSATSPERRDVIIIVSGIYGLSGPEEREMVVS LRTMEIERNELRLKVLVDIQARNDIDRGTFRRGVVEIFPASPDEHCYRVFPG DEIERIREVDALTGSLIDRDHVALFPASHVTRAEKKEKAIQNTLEKLEBOLKYMHE NGKLEAQRLEQRTYVDLEMMREMGFCGIEYNSHLLTRPGSTPYLLDYFPDPM IVDSSHVTLIPQVRGMFGDQARKQVLVDHGRFLPSALDNPRPREFEKKHNTVYV SATPGVRIEHTDMEVBOIIRPTGLDLIDVRPIEGQIDPLIGRIQARIEBERVAV TLTTRKMBEDLDYIKKEIGIKVNYHSRILKTLERELRDLRIGKYDVLVGINLRG LDIPEVSLVALIDADKRGFLNSERSLITOTIGRAANAGRVIMYADKTKSMELINE TKRREQOERFEEHGIPTKTIKEIRIVIRVTAABDGAETKTAAPGLSMTKKER OKVVEQMEHMEKKAALDFERAEALRDLLELKAEG"
complement (5166..5396)
/gene="csbA"
/locus_tag="BSU35180"
complement (5166..5396)
/gene="csbA"
/locus_tag="BSU35180"
/function="unknown"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAB15535.1"
/db_xref="GI:2636044"
/db_xref="GOA:P37953"
/db_xref="Subtilist:BG10501"
/db_xref="UniProt/Swiss-Prot:P37953"
/translation="MITKAVFALFPPMLVLFTRVTENHYVALTAAALPASYLKG YETLYFVIGLDVAVSLVAGGLYMAKKAERKEE"
5843..8338
/gene="yvkK"
/locus_tag="BSU35190"
5843..8338
/gene="yvkK"
/locus_tag="BSU35190"
/function="unknown"
/note="similar to pyruvate, water dikinase"
/codon_start=1
/transl_table=11
/protein_id="CAB15536.1"
/db_xref="GI:2636045"
/db_xref="GOA:O34796"
/db_xref="InterPro:IPR002192"
/db_xref="InterPro:IPR008279"
/db_xref="Subtilist:BG14114"
/db_xref="UniProt/TrEMBL:O34796"
/translation="MYSVLFQAEBSQLAGKGMNLIKLTRGLPVPDGFIIQTNAL ARMENNOERSENVEGGLIGTFSDLELDELTSFYLRSSYNSAVRSSASBDL BGASFGQYETELNITKTEREPLAKYSCASFPFGSSYKKKNNQIAPLMGIYVO GLIDSMGCVITISRNVTYDDRELLISAAYCGEAVVSGNVTPTPIYKSSREIIOE IGAKETIMESAAEGIAEKETSEDMKSRFCLTBQVLELAEITKTEDLYGYPIIDIEG IADHQIYLLQARPITITIDDKAAAEKRSFMTTDMDFPMLMESNISIGPVSPLTSE FIVPALEYGKSKSMQKFPIGVVVDEVLKRGHITYSNQGQGPSPSDCKEKLPIISE HWYDIINHTYLPYRTLDQAOETHEAASALEFQGLKAFAYLTAVEHFNIVPQIIL TNLQAMYDIOGESNAHFYELTCKNNKSLSETDRCLMLFSVEVQENPLALAIPENN

gene

CDS

```

      KPEQLQEKLEQETDEGRHFLKNVHEFLQEYGMRSVKSHDLLEQIWENPEYFALANIQNY
      VRNGYHDFNFEQFKTEKREKLYNEFLSEIDEPGRTEREDRYQATLNSANTKDHHFY
      IDALDKARIFLTKIGLELAENGVIQIORDLWFLYDEVOALHPHSLOEKREKR
      OIFHEVELAQA PAYLGPTEKOLKAEBEIVAVI EDEKNTENH IEGIAASGIATGPV
      KIIRDANFSPAPRGDVI VCKMTTPLMTSLERQDAKAIITTDGILSHAIITAREYGIP
      AVLGRTATERTLRDGDIIITVDGSSGKITVRS"
      gene
      8141..8983
      /gene="yvkB"
      /locus_tag="BSU35200"
      8414..8983
      /gene="yvkB"
      /locus_tag="BSU35200"
      /function="unknown"
      /note="similar to transcriptional regulator (Tetr/AcrR
      family)"
      /codon_start=1
      /transl_table=11
      /protein_id="CAB15537.1"
      /db_xref="GI:2636046"
      /db_xref="GOA:O34643"
      /db_xref="InterPro:IPR001647"
      /db_xref="Subtilist:BG14113"
      /db_xref="UniProt/TREMBL:O34643"
      Query Match      98.7%; Score 1166.4; DB 1; Length 200690;
      Best Local Similarity 99.5%; Pred.No.2.6e-249;
      Matches 1170; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
      QY      7 TGGTTACTCATTTATAGCCTTGCTGCTCTACTCTGTCATCGGAATTTTGAAAAACAGACA 66
      DB      90671 TGGTTACTCATTTATAGCCTTGCTGCTCTACTCTGTCATCGGAATTTTGAAAAACAGACA 90612
      QY      67 CATCGAAGAAAAATTGATGATGCCCTCCCTGTCGGGGGATATTAAAGGATCGCGGAGAAA 126
      DB      90611 CATCGAAGAAAAATTGATGATGCCCTCCCTGTCGGGGGATATTAAAGGATCGCGGAGAAA 90552
      QY      127 TCGACTGTGACAGAGCGTCAACACCGGAATATTATAGAACCGCGTTACAGAAGCTGTGGA 186
      DB      90551 TCGACTGTGACAGAGCGTCAACACCGGAATATTATAGAACCGCGTTACAGAAGCTGTGGA 90492
      QY      187 AAAACACACGAGACAGATGTCAGAAATGATTTACTGTGGACACACCGGAGAGAAAAAGCCGATT 246
      DB      90491 AAAACACACGAGACAGATGTCAGAAATGATTTACTGTGGACACACCGGAGAGAAAAAGCCGATT 90432
      QY      247 AAAACGGAACCTCAGAGGGCCGGAATATCGAGACAAAAAGAAAGTATGATGAGAAACAGTA 306
      DB      90431 AAAACGGAACCTCAGAGGGCCGGAATATCGAGACAAAAAGAAAGTATGAGAAACAGTA 90372
      QY      307 GAAAGAGGGGCTAACGCGATTGTCAAGTATGATGCGTGTTTAACCCAGATTATCAAAATC 366
      DB      90371 GAAAGAGGGGCTAACGCGATTGTCAAGTATGATGCGTGTTTAACCCAGATTATCAAAATC 90312
      QY      367 ATCTTTTCGAGAAAGCTTCTGCGAGGCCAAATATCGCGCTCATTTGTGAATGTTTAGAAGAC 426
      DB      90311 ATCTTTTCGAGAAAGCTTCTGCGAGGCCAAATATCGCGCTCATTTGTGAATGTTTAGAAGAC 90252
      QY      427 CATATGATGATCATGGGGCCGCGCTTATGATGAATTTGAGAGAGGCTTTACCGCTCAATTT 486
      DB      90251 CATATGATGATCATGGGGCCGCGCTTATGATGAATTTGAGAGAGGCTTTACCGCTCAATTT 90192
      QY      487 CCTTATATATGCGCATCTTGTCAATTACAGATAGTGAATATACCGAGTTCTTTAAACAAAA 546
      DB      90191 CCTTATATATGCGCATCTTGTCAATTACAGATAGTGAATATACCGAGTTCTTTAAACAAAA 90132
      QY      547 GCAAAAGAACGAAACACAAAGTCAATCTGCTGATTAACCTCAAAATTTACAGATAGATAT 606
      DB      90131 GCAAAAGAACGAAACACAAAGTCAATCTGCTGATTAACCTCAAAATTTACAGATAGATAT 90072
      QY      607 TTAGCTATTTTGAATACATGATTTCCCTGTATACGCTTCTGCGCGTGGGCTGGCT 666
      DB      90071 TTAGCTATTTTGAATACATGATTTCCCTGTATACGCTTCTGCGCGTGGGCTGGCT 90012
      QY      667 CAAGCACTCGGCATTGACGAAGAAAACAGCATTTAAGGAATGCTGAATGCGCGCCAGAT 726
```

```

      DB      90011 CAAGCACTCGGCATTGACGAAGAAAACGCAATTAAGGAAATGCGAATGCGCCCGCAT 89952
      QY      727 CCGGAGACCAATGAGAAATTTCTCGCTGATCAAGTCCGAGCGAGCTGGGCACTTTGTTAAT 786
      DB      89951 CCGGAGACCAATGAGAAATTTCTCGCTGATCAAGTCCGAGCGAGCTGGGCACTTTGTTAAT 89892
      QY      787 GGGTTTCCCGCAACGACGCTTCTTCTTCACTTGAATATATGAAAACTGTAAAAAATATC 846
      DB      89891 GGGTTTCCCGCAACGACGCTTCTTCTTCACTTGAATATATGAAAACTGTAAAAAATATC 89832
      QY      847 GGTATACCGACCAATGATTCGATCATCATGAACTGCCGCGACACCGGTGATTCGG 906
      DB      89831 GGTATACCGACCAATGATTCGATCATCATGAACTGCCGCGACACCGGTGATTCGG 89772
      QY      907 ACACAGCAATTCGCAATGACAGTATTCGCTTATATTGAAGCAAGTGAATGATCTTAATC 966
      DB      89771 ACACAGCAATTCGCAATGACAGTATTCGCTTATATTGAAGCAAGTGAATGATCTTAATC 89712
      QY      967 GGTGAAACACAGAACCGATCGTAAAAAGCCTATGAGAGAGCAAAATTCCTGCAGACAA 1026
      DB      89711 GGTGAAACACAGAACCGATCGTAAAAAGCCTATGAGAGAGCAAAATTCCTGCAGACAA 89652
      QY      1027 CTGCATGACCTAGAGTATTAAGTCAACAGATGAATTTTGAATTTTGAAGAAAAAGATG 1086
      DB      89651 CTGCATGACCTAGAGTATTAAGTCAACAGATGAATTTTGAATTTTGAAGAAAAAGATG 89592
      QY      1087 CACAAACGGTGCATATATATGAGTGGCGGCAATTTATCATGAGTGGCGGACCTTTAATGAA 1146
      DB      89591 CACAAACGGTGCATATATATGAGTGGCGGCAATTTATCATGAGTGGCGGACCTTTAATGAA 89532
      QY      1147 AAAATCCACGAATACAAAGTAAAGACAGCTCGTAAGC 1182
      DB      89531 AAAATCCACGAATACAAAGTAAAGACAGCTCGTAAGC 89496
```

```

      RESULT 7
      AB046355
      LOCUS
      DEFINITION
      ACCESSION
      VERSION
      KEYWORDS
      SOURCE
      ORGANISM
      Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
      REFERENCES
      1 Urushibata,Y., Tokuyama,S. and Tahara,Y.
      Characterization of the Bacillus subtilis ywsc gene, involved in
      gamma-polyglutamic acid production
      JOURNAL
      PUBMED
      11751809
      J. Bacteriol. 184 (2), 337-343 (2002)
      2 (bases 1 to 3738)
      Tahara,Y. and Urushibata,Y.
      Direct Submission
      Submitted (18-JUL-2000) Yasutaka Tahara, Shizuoka University,
      Faculty of Agriculture, 836 Ohya, Shizuoka city, Shizuoka 422-8529,
      Japan (E-mail:acytaha@agr.shizuoka.ac.jp,
      Tel:81-54-238-4878(fax.7808), Fax:81-54-237-3028)
      FEATURES
      source
      1..3738
      /organism="Bacillus subtilis"
      /mol_type="Genomic DNA"
      /db_xref="taxon:1423"
      497..1678
      /gene="ywsc"
      497..1678
      /gene="ywsc"
      /codon_start=1
      /transl_table=11
      /protein_id="BAB40948.1"
      /db_xref="GI:13591557"
      /translation="MWLLITACAVILVIGILEKRRHQKIDALPVRVNINGIRGKSTV
```

```

TRLTGTLIENGAKTKTGCTGTDARMTYMDPEKPKIRKROGPNIGEOKEVMEIYE
RGAATVISECMANAVPDIOIIFOEELQANIVIVNVEDHDWNGPFLIDEIABEPTAT
IPNGHLVITDSEYTERFKQAKERNTKVIIADNSKITDEYLRFBETWPFDPNLSLML
GVAQALGIDESTAFKMLNAPDPGAKRIILPLISPSBGFHPNFPANDASTLNTIK
RKVEIGYPTDDPIIIMNCRADRVDTQOFADVLPLYEASLILIGEOEPIVAVYER
GKIIPADKLHDEYKSTDEIMELKKRHNRIYIGVGNHGAEPLEIKIHEYKVKQIV
S"
gene
1693..2142
/gene="ywcA"
CDS
1693..2142
/gene="ywcA"
/codon_start=1
/transl_table=11
/protein_id="BAB40949.1"
/db_xref="GI:13591558"
/translation="MFSGDYLIALILGLVLSLIPAKETGIVPAGLVNPGYGLVFNOP
VEILLIVLSLLIVIVKYGSKEMILYGRKRFAMLTGLIVLKIAPFLPIVVEFI
AEPRGIGIIVPGLIANTIQKQGLTTFGSTILLSGRAPAIMFVYLLI"
2161..3303
/gene="ywcB"
CDS
2161..3303
/gene="ywcB"
/codon_start=1
/transl_table=11
/protein_id="BAB40950.1"
/db_xref="GI:13591559"
/translation="MKKELSPHEKLKLTQOKKKTKNKHVFIPIVFLVFMFAEWAG
KAETPKVATSDVLSASFVGDIMMGRVVEKTEQKGDLSIFQYVEFIRASDVYAGN
FENPVTQKNYKQADKEIHLQTNKESVVLKDMNFTVNSANNHMDYGVQGMKDTLG
EAPATKLDIVAGVATLSPAKRKISYQKNGVTLATLGFYDSGKGFAPKPTPGVLP
DPEITPMISEKAKRADIVVOSHNGEYDNDPRNQQLARMSDADADIIIVGHHP
VLEPLEVNGVITFIYSLGNFVFDQMTFRDSALVQYHLKKNTERFEVTPIDIHET
PAPVKDLSLKQTIIRLETKDSNFAMKVEDKLTPIIDHSDLSKX"
3318..3485
/gene="ywcC"
CDS
3318..3485
/gene="ywcC"
/codon_start=1
/transl_table=11
/protein_id="BAB40951.1"
/db_xref="GI:13591560"
/translation="MKPVKAIWPFVAVALVFMFMSAFKFNQDLTDEQKQIDIMENKXI
QOQEEPVANK"

```

ORIGIN

Query Match 98.5%; Score 1164.8; DB 1; Length 3738;
 Best Local Similarity 99.4%; Pred. No. 1,1e-248;
 Matches 1169; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

OY 7 TGGTACTCATTAATAGCCGTGTCTGTCAATCTGTCATCGGAATATTAGAAAAACGACGA 66
DB TGGTACTCATTAATAGCCGTGTCTGTCAATCTGTCATCGGAATATTAGAAAAACGACGA 559
OY 67 CATCGAAAAACATTGATGCCCTCCTCTGTTGGGTGAATATTAAACGGCATCCGCCGAAAA 126
DB CATCGAAAAACATTGATGCCCTCCTCTGTTGGGTGAATATTAAACGGCATCCGCCGAAAA 619
OY 560 CATCGAAAAACATTGATGCCCTCCTCTGTTGGGTGAATATTAAACGGCATCCGCCGAAAA 619
DB TCGACTGTGACAGGCTGACACCCGGAATATTAAAGCCGGTTACAAAGATGTTTGA 186
DB TCGACTGTGACAGGCTGACACCCGGAATATTAAAGCCGGTTACAAAGATGTTTGA 679
OY 187 AAAACAACAGAACAGATCAAGAAATGATTACTGGGACACACCGGAGGAAAAAGCCGATT 246
DB AAAACAACAGAACAGATCAAGAAATGATTACTGGGACACACCGGAGGAAAAAGCCGATT 739
OY 247 AAAACGAAACCTTACGGGGCCGAATATTCGAGAGCAAAAAAGAACTCATGAGAGAAACAGTA 306
DB AAAACGAAACCTTACGGGGCCGAATATTCGAGAGCAAAAAAGAACTCATGAGAGAAACAGTA 799
OY 307 GAAAGAGGGGCTAACGGAATTGTCAATGATGCAATGGATGGCTTTAACCCAGATTATCAAAATC 366
DB GAAAGAGGGGCTAACGGAATTGTCAATGATGCAATGGATGGCTTTAACCCAGATTATCAAAATC 859

```

```

OY 367 ATCTTCAGAGAAAGAACTTCTGACAGGCCAATATCGGCTCATTTGTAATGTTTGAAGAC 426
DB ATCTTCAGAGAAAGAACTTCTGACAGGCCAATATCGGCTCATTTGTAATGTTTGAAGAC 919
OY 427 CATATGATGTCATGAGGCGCGACGCTTGATGAATTCGAGAAAGGCTTTACCGCTACAAAT 486
DB CATATGATGTCATGAGGCGCGACGCTTGATGAATTCGAGAAAGGCTTTACCGCTACAAAT 979
OY 487 CCTTATAATGGCCATCTTTCATCATACAGATGATGATATACCGAGTTCTTTAAACAAAA 546
DB CCTTATAATGGCCATCTTTCATCATACAGATGATGATATACCGAGTTCTTTAAACAAAA 1039
OY 547 GCAAAAGAACGAAACACAAAAGTCATCATCTGCTATTAATCAAAAATTTACAGATGAGTAT 606
DB GCAAAAGAACGAAACACAAAAGTCATCATCTGCTATTAATCAAAAATTTACAGATGAGTAT 1039
OY 1040 GCAAAAGAACGAAACACAAAAGTCATCATCTGCTATTAATCAAAAATTTACAGATGAGTAT 1039
DB TTAGCTTAATTTGAATATCATGATATTCCTGATTAACGCTTCTGCGCTGGGTGGCT 1159
OY 607 TTAGCTTAATTTGAATATCATGATATTCCTGATTAACGCTTCTGCGCTGGGTGGCT 666
DB TTAGCTTAATTTGAATATCATGATATTCCTGATTAACGCTTCTGCGCTGGGTGGCT 1159
OY 667 CAAGCACTCGGCATTTGACGAGAAACAGCATTTAAGGAAATGCTGAATGCCGCCAGAT 726
DB CAAGCACTCGGCATTTGACGAGAAACAGCATTTAAGGAAATGCTGAATGCCGCCAGAT 1219
OY 1160 CAAGCACTCGGCATTTGACGAGAAACAGCATTTAAGGAAATGCTGAATGCCGCCAGAT 1219
DB CCGGAGCAATGAGAAATTTCTCCGCTGATCAGTCCGACGACCTGGGCACTTTGTTAAT 786
OY 727 CCGGAGCAATGAGAAATTTCTCCGCTGATCAGTCCGACGACCTGGGCACTTTGTTAAT 786
DB CCGGAGCAATGAGAAATTTCTCCGCTGATCAGTCCGACGACCTGGGCACTTTGTTAAT 1279
OY 787 GGGTTTGGCGGAAACGACGCTTTCTTCTATCTTTGAATATATGAAAACGTTTAAAGAAATC 846
DB GGGTTTGGCGGAAACGACGCTTTCTTCTATCTTTGAATATATGAAAACGTTTAAAGAAATC 1339
OY 1280 GGGTTTGGCGGAAACGACGCTTTCTTCTATCTTTGAATATATGAAAACGTTTAAAGAAATC 1339
DB GGGTTTGGCGGAAACGACGCTTTCTTCTATCTTTGAATATATGAAAACGTTTAAAGAAATC 1339
OY 847 GGTATCCCGACCGATGATCCGATCATCATGAACTGCCCGGACGAGCGGTGATGCG 906
DB GGTATCCCGACCGATGATCCGATCATCATGAACTGCCCGGACGAGCGGTGATGCG 1399
OY 1340 GGTATCCCGACCGATGATCCGATCATCATGAACTGCCCGGACGAGCGGTGATGCG 1399
DB GGTATCCCGACCGATGATCCGATCATCATGAACTGCCCGGACGAGCGGTGATGCG 1399
OY 907 ACACAGCAATTCGCAATGAGATATGCTTATATTTAGAGCAAGTGAATGATCTTAATC 966
DB ACACAGCAATTCGCAATGAGATATGCTTATATTTAGAGCAAGTGAATGATCTTAATC 1459
OY 1400 ACACAGCAATTCGCAATGAGATATGCTTATATTTAGAGCAAGTGAATGATCTTAATC 1459
DB GGTGAAACAAACGACCGATGTAAGACCTTATGAAAGCAAAATTCCTGCAACAAA 1519
OY 967 GGTGAAACAAACGACCGATGTAAGACCTTATGAAAGCAAAATTCCTGCAACAAA 1026
DB GGTGAAACAAACGACCGATGTAAGACCTTATGAAAGCAAAATTCCTGCAACAAA 1519
OY 1027 CTGCATGACCTTAGATATTAAGTCAACAGATGAATTTATGAAATGTTTAAAGAAAGATG 1086
DB CTGCATGACCTTAGATATTAAGTCAACAGATGAATTTATGAAATGTTTAAAGAAAGATG 1579
OY 1520 CTGCATGACCTTAGATATTAAGTCAACAGATGAATTTATGAAATGTTTAAAGAAAGATG 1579
DB CACAACCGGTGATATATGAGGTGGGCAATATTTCAATGTTGCGGACAGACCTTTAATTGA 1146
OY 1087 CACAACCGGTGATATATGAGGTGGGCAATATTTCAATGTTGCGGACAGACCTTTAATTGA 1146
DB CACAACCGGTGATATATGAGGTGGGCAATATTTCAATGTTGCGGACAGACCTTTAATTGA 1639
OY 1580 CACAACCGGTGATATATGAGGTGGGCAATATTTCAATGTTGCGGACAGACCTTTAATTGA 1639
DB AAAATTCACGAATACAGGTTAAGCAGGCTCGTAAC 1182
OY 1147 AAAATTCACGAATACAGGTTAAGCAGGCTCGTAAC 1182
DB AAAATTCACGAATACAGGTTAAGCAGGCTCGTAAC 1675

```

RESULT 8
 BS292954 8189 bp DNA linear BCT 02-DEC-2004
 locus
 DEFINITION B. subtilis ywe(A,B,C,D,E,F,G) and gerBC genes.
 ACCESSION Z92954
 VERSION Z92954.1 GI:1894764
 KEYWORDS gerBC gene; ywcC gene; ywcA gene; ywcB gene; ywcD gene;
 ywcE gene; ywcF gene; ywcG gene.
 SOURCE Bacillus subtilis
 ORGANISM Bacillus subtilis
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 1 (bases 1 to 8189)
 PRESCAN,E., Moszer,I., Boursier,L., Cruz Ramos,H.C., de la
 Fuente,V., Hullo,M.F., Lelong,C., Schleich,S., Sekowska,A.,
 Song,B.H., Villani,G., Kunst,F., Danchin,A. and Glaser,P.
 The Bacillus subtilis genome from gerBC (311 degrees) to lscr (334

degrees)
JOURNAL Microbiology (Reading, Engl.) 143 (PT 10), 3313-3328 (1997)
PUBMED 9353933
REFERENCE 2 (bases 1 to 8189)
AUTHORS LeJong, C., Glaser, P., Pressecan, E. and Danchin, A.
TITLE Bacillus subtilis gerb downstream sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 8189)
AUTHORS Glaser, P.
TITLE Direct Submission
Submitted (13-MAR-1997) Philippe Glaser, Regulation de l'Expression
Genetique, Institut Pasteur, 28 Rue du Dr Roux, Paris, 75724,
FRANCE
FEATURES
source Location/Qualifiers
1. 8189
/organism="Bacillus subtilis"
/mol_type="genomic DNA"
/strain="168"
/db_xref="taxon:1423"
1. 1069
/gene="ywcC"
/gene="ywcC"
/gene="ywcC"
/note="product highly similar to Bacillus anthracis CapB
protein"
/codon_start=2
/transl_table=1
/protein_id="CAB07475.1"
/db_xref="GI:1894765"
/db_xref="GOA:Q798J2"
/db_xref="UniProt/TREMBL:Q798J2"
/translation="RGKSTVRLTGTILIEAGYKTVGKTGTGDARMIYMDPEEPRK
RKQPGNIGGEQKWEKRETVBERGANATVSECVANVPDQILFQEEILQANIGVIVNVE
DHMDVMEPTLDELAEATPATTPVNGHLVITDSETEPFQAKERAKVITADSKIT
DEVIYRKEKVMFPDNASIALGVAGALGIDETATKGMINAPDGMKILPLISPSBP
GHFVNGFANASSTLINIKRKGKIGYPTDPIIIMNCRADRVDTQOFADVLPYLE
ASELILGETEPIVKAIVEGKIIPADKLHDEYKSTDEIMELKKSMHNRVITYGVNII
HGAEPLIEKHEKVKOLVS"
join(1069..1075,1084..1533)
/gene="ywcA"
1069..1075
/gene="ywcA"
1084..1533
/gene="ywcA"
/note="product highly similar to Bacillus anthracis CapC
protein"
/codon_start=1
/transl_table=1
/protein_id="CAB07468.1"
/db_xref="GI:1894766"
/db_xref="GOA:P96737"
/db_xref="UniProt/TREMBL:P96737"
/translation="WFGSDLYIALILGVLSLIPAEKTIIVPAGLVPGVGLVNOP
VFILLVAVLSLLTVIVYKGLSKFMILYGRKRFAMLITGLVLIAPDFLPIVFPFI
AEFGIGIIVPLIANTIOKQGLITTFGSTILLSGATPAIMFVYLLI"
1538..2694
/gene="ywcB"
1538..1544
/gene="ywcB"
1552..2694
/gene="ywcB"
/note="product highly similar to Bacillus anthracis CapA
protein"
/codon_start=1
/transl_table=1
/protein_id="CAB07469.1"
/db_xref="GI:1894767"
/db_xref="UniProt/TREMBL:P96738"
/translation="MKKELSFHEKLIKLTQKQKKTKNGVIAIPVIVVLMFAFMKAC
KAETPKATISDVVASFVSDIMNGRVEVETQKGDSTIFQYVEPIFRASDVVAG
FENVTYQKNYKQADKEIHLQTNKESVYVLDNMFTVINSANNHAMDDVGVMKDYLQ
EFAQONLDIVAGYSLDAKKRISQKNGVTIATLFTGTVSGKGFAPAKKNTPGVLP
DPEIFIPWISAKKHADIIVVQSHMGQYDNDPNDROQLARAMSADAGADIIVGHNPH

gene
2697..2914
/gene="ywcC"
2697..2702
/gene="ywcC"
2709..2876
/gene="ywcC"
/codon_start=1
/transl_table=1
/product="unknown"
/protein_id="CAB07470.1"
/db_xref="GI:1894768"
/db_xref="UniProt/Swiss-Prot:P96739"
/translation="MKFVKAIWPFVAVAIWFMMSAFKFNDDLTDEKQKIDMENKI
QOGEFVNANK"
2880..2914
/gene="ywcC"
3000..4258
/gene="ywcD"
3000..3007
/gene="ywcD"
3017..4258
/gene="ywcD"
/note="product similar to E. coli NUPC protein and to
Listeria species p60-related proteins"
/codon_start=1
/transl_table=1
/protein_id="CAB07476.1"
/db_xref="GI:1894769"
/db_xref="UniProt/Swiss-Prot:P96740"
/translation="NNTLAWKKELVAVIICFLVPIPMTKAIEADTSSSELIYSEAK
NLGQYKYGKGEKPKGPDGSGIYVFSKADILHPSVNDQYKIGTAVNPKPD
ILFPKSGSGTVPHTDALYIGQGMHSTQSKGVITTYKKSSYNGGTIGARRIA
DPAITADVPVQAEKRTYGVYVFGSTPSGFCGLVQVVFQALGIVLRSEQDM
AVGKQVAPQMIKKGDDVYFENYTKTGISHAGIYAGAGRFTQASRSEKVTISYSEDY
KSKWTKIRFDNLTIPKKNPISBATLYGVEVYKQGVTPETGFDAGVQVYVQYA
AGISLPYATSOYNAGTKIKKADLPEDIVFPQSTLNPSTIYIGNGVAVHTLNGVT
ITNMNTSTYKDKKXGIRVQ"
4259..4293
complement(4292..5170)
/gene="ywcE"
complement(4292..5152)
/gene="ywcE"
/note="product similar to Bacillus subtilis YxeH and YxeS
proteins and to E. coli Yida protein"
/codon_start=1
/transl_table=1
/protein_id="CAB07471.1"
/db_xref="GI:1894770"
/db_xref="GOA:P96741"
/db_xref="UniProt/TREMBL:P96741"
/translation="MKCTAIDLDGTLINKSVISAENREAIKRAVDAGILVTICTGNA
TFVDKALDDLDLPIIAANGTIDHTGRILISLMDQEGKAIADLLSKNIYFEVY
TDHILSPFDGEAKLHAELDILKSANNEQTDLMQGMATQFGFGIKPIPIHESVD
GGENIYKLCFSPMDKLOAKELKHKKLAOTSSGKHIIELIPASGGALATKA
DIVIGIERODIYATGDSPNDSMFVAGHRIAMENALDELKESFTFYKSDENGVAIF
IDQLLSQYA"
complement(5163..5170)
/gene="ywcE"
5277..5305
/gene="ywcF"
5277..5305
/gene="ywcF"
complement(5308..6291)
/gene="ywcF"
/pseudo
complement(5308..6276)
/gene="ywcF"
/note="an in frame TGA stop codon is found at position
5599 to 5601
translation similar to Bacillus subtilis LyxR protein"

```

/psedu
/codon_start=1
/transl_table=1
/complement (6285, .6291)
/gene="ywtf"
6594, .8112
/gene="ywtf"
6594, .6601
/gene="ywtf"
6609, .7982
/gene="ywtf"
/ncore="product highly similar to metabolite transport
protein"
/codon_start=1
/transl_table=1
/protein_id="CAB07473.1"
/db_xref="GI:1894771"
/db_xref="GOA:P96742"
/db_xref="UniProt/TREMBL:P96742"
/transl_table=1
/translation="MKKOSNIMLYFFGALGALGYDTGVISGAILFHKELGLNAFT
EGLVSSILVGAIISSGAGKLTDRPKKAIMAALFICIGLGVALAPMGVAF
RIILGAVGTTTIVLYSELAPKHKALSLNQLMTVIGILSYIVNYIPADAE
MRWMLGAAVPSLILIGILFMPSPRMLPTNGESKAKKILEKLRGKIDOEIHD
KEAKODBGKLELPDMPVPRALLAGLLOPITNTIYYAAPTFTVGRNSA
SILGVIGIVNMLTVAIKIIDKIGKPLIFGNAGMVISLYLALVNLFPNTPA
ASMTVVICLGVIFVAVSWGPVWMLPELPLHRSIGVSTLMLHVLIVSLT
YPIIMEAIGISYLFILYNAIGIMAFVFRFVETKRSLEIEDLRNKGQAG
KOOTVGT"
7981, .8112
/gene="ywtf"
/complement (8019, .8189)
/gene="gerbc"
/complement (8019, .>8189)
/gene="gerbc"
/codon_start=1
/transl_table=1
/product="GerBC"
/protein_id="CAB07474.1"
/db_xref="GI:1894772"
/db_xref="GOA:P96743"

terminator
7981, .8112
/gene="ywtf"
/complement (8019, .8189)
/gene="gerbc"
/complement (8019, .>8189)
/gene="gerbc"
/codon_start=1
/transl_table=1
/product="GerBC"
/protein_id="CAB07474.1"
/db_xref="GI:1894772"
/db_xref="GOA:P96743"

Query Match 89.4%; Score 1056.4; DB 1; Length 8189;
Best Local Similarity 99.4%; Pred. No. 1.3e-224;
Matches 1060; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

```

537 TAAACAAAAGACAAAGAACGAAACACAAAAGTCATCTGCTGATACTCAAAATTAC 596
421 TAAACAAAAGACAAAGAACGAAACACAAAAGTCATCTGCTGATACTCAAAATTAC 480
597 AGATGATATTACGTAATTTTGAATCATGATATTCCTGATTAACGCTTCTGGCGCT 656
481 AGATGATATTACGTAATTTTGAATCATGATATTCCTGATTAACGCTTCTGGCGCT 540
657 GGGTGTGCTCAACACCTCGGCAATTTGAGAAAGACATTTAAGGGAATGCTGAATGC 716
541 GGGTGTGCTCAACACCTCGGCAATTTGAGAAAGACATTTAAGGGAATGCTGAATGC 600
717 GCCCCAGATCCGGAGCAATGAGAAATTCCTGCTGATCAGTCGAGCGAGCTGGGCA 776
601 GCCCCAGATCCGGAGCAATGAGAAATTCCTGCTGATCAGTCGAGCGAGCTGGGCA 660
777 CTTTGTAAATGGGTTTGGCGAAACGACGCTTCTTCACTTTGAATATATGAAACGCT 836
661 CTTTGTAAATGGGTTTGGCGAAACGACGCTTCTTCACTTTGAATATATGAAACGCT 720
837 AAAAGAAATCGGTTACCCGACCCGATGATCCGATCATCATGAACTGGCGAGACCG 896
721 AAAAGAAATCGGTTACCCGACCCGATGATCCGATCATCATGAACTGGCGAGACCG 780
897 TGTGATCGGACACAGCAATTCGCAATGACGTAATGCTTATATTAAGCAATGAACT 956
781 TGTGATCGGACACAGCAATTCGCAATGACGTAATGCTTATATTAAGCAATGAACT 840
957 GATCTTAATCGGTTAAACACAGAAACCGATCGTAAGCCTATGAGAGCAAAATTC 1016
841 GATCTTAATCGGTTAAACACAGAAACCGATCGTAAGCCTATGAGAGCAAAATTC 900
1017 TGCAGACAACTGCACTGACCTTGAATTAATGCAACATTAATTAAGAAATTTTAA 1076
901 TGCAGACAACTGCACTGACCTTGAATTAATGCAACATTAATTAAGAAATTTTAA 960
1077 GAAAGAAATGCAACCGGTCATATATGCGTGGCAATATTCAGTGGCGAGAGCC 1136
961 GAAAGAAATGCAACCGGTCATATATGCGTGGCAATATTCAGTGGCGAGAGCC 1020
1137 TTTAATGAAAAAATCCAGATACAGATTAAGCAAGCTCGTAAC 1182
1021 TTTAATGAAAAAATCCAGATACAGATTAAGCAAGCTCGTAAC 1066

RESULT 9
AB017333 36/c
WPCOMMENT
Sequence split into 43 fragments LOCUS AB017333 Accession AB017333
Fragment Name Begin End
AB017333_00 1 110000
AB017333_01 100001 210000
AB017333_02 200001 310000
AB017333_03 300001 410000
AB017333_04 400001 510000
AB017333_05 500001 610000
AB017333_06 600001 710000
AB017333_07 700001 810000
AB017333_08 800001 910000
AB017333_09 900001 1010000
AB017333_10 1000001 1110000
AB017333_11 1100001 1210000
AB017333_12 1200001 1310000
AB017333_13 1300001 1410000
AB017333_14 1400001 1510000
AB017333_15 1500001 1610000
AB017333_16 1600001 1710000
AB017333_17 1700001 1810000
AB017333_18 1800001 1910000
AB017333_19 1900001 2010000
AB017333_20 2000001 2110000
AB017333_21 2100001 2210000
AB017333_22 2200001 2310000

```

AE017333_23 2300001 2410000
AE017333_24 2400001 2510000
AE017333_25 2500001 2610000
AE017333_26 2600001 2710000
AE017333_27 2700001 2810000
AE017333_28 2800001 2910000
AE017333_29 2900001 3010000
AE017333_30 3000001 3110000
AE017333_31 3100001 3210000
AE017333_32 3200001 3310000
AE017333_33 3300001 3410000
AE017333_34 3400001 3510000
AE017333_35 3500001 3610000
AE017333_36 3600001 3710000
AE017333_37 3700001 3810000
AE017333_38 3800001 3910000
AE017333_39 3900001 4010000
AE017333_40 4000001 4110000
AE017333_41 4100001 4210000
AE017333_42 4200001 4222645
Continuation (37 of 43) of AE017333 from base 3600001 (AE017333 Bacillus licheniformis)

Query Match 65.4%; Score 772.8; DB 1; Length 110000;
Best Local Similarity 78.6%; Pred. No. 1.2e-161;
Matches 924; Conservative 0; Mismatches 252; Indels 0; Gaps 0;

QY 7 TGGTATCATATTAGCCTGTGCTGTCTACTGTGCATCGGAATATTGAAAAAGACGA 66
Db 67582 TGGTATCATATTAGCCTGTGCTGTCTACTGTGCATCGGAATATTGAAAAAGACGC 67523
QY 67 CATCAGAAAAATTGATGCTCCCTGTCGGGTGTAATTAAACGCATCCGGGAAA 126
Db 67522 CACCGAGAAAATTGATGCTCCCTGTCGGGTGTAATTAAACGCATCCGGGAAA 67463
QY 127 TCGACTGTGACAAAGCTGACAAACCGGAATATTAGAACCGGTTCACAACTGTTGA 186
Db 67462 TCCACGGTGACAAAGCTGACAAACCGGAATATTAGAACCGGTTCACAACTGTTGA 67403
QY 187 AAAACAACAGAACGATGACAAAGATGTTTACTGGGACACACCGGAGAAAGCCGATT 246
Db 67402 AAAACAACCGGACGACGACGAAAGATGTTTACTGGGACACACCGGAGAAAGCCGATT 67343
QY 247 AAACGGAAACCTCAGGGGCCGAATATCGAGAGCAAAAAAGAGTCATGAGAAAAACATA 306
Db 67342 AAAAGAAACCGCAGAGGGCCGAATATCGAGAGCAAAAAAGAGTCATGAGAAAAACATA 67283
QY 307 GAAAGAGGGGCTAACGCGATTGTCACTGTAATGATGCTGTTAACCCGATTATCAATC 366
Db 67282 GAAAGAGGGGCTAACGCGATTGTCACTGTAATGATGCTGTTAACCCGATTATCAATC 67223
QY 367 ATCTTTCAGAAAGACTTCGAGGCCAATATCGGCGCTATTGTGAATGTTTGAAGAC 426
Db 67222 ATCTTTCAGAAAGACTTCGAGGCCAATATCGGCGCTATTGTGAATGTTTGAAGAC 67163
QY 427 CATATGATGTCAATGAGGCCGACGCTTGATGAAATTGCAAGAGCTTTACCGCTACATT 486
Db 67162 CATATGATGTCAATGAGGCCGACGCTTGATGAAATTGCAAGAGCTTTACCGCTACATT 67103
QY 487 CCTTATAATGGCCATCTTGTCAATACAGATAGTGAATATACCGAGTTCTTTAAACAAAA 546
Db 67102 CCTTATAATGGCCATCTTGTCAATACAGATAGTGAATATACCGAGTTCTTTAAACAAAA 67043
QY 547 GCAAAAGAGAAACAAAGATCATCTGTGTAATCAAAAAATTACAGATGAGTAT 606
Db 67042 GCAAAAGAGAAACAAAGATCATCTGTGTAATCAAAAAATTACAGATGAGTAT 66983
QY 607 TTACGTAATTTGTAATACATGTAATCCCTGATTAACGCTTCTGCGCTGGGTGTGCT 666
Db 66982 CTGAGACAGTTTGAATACATGTAATCCCTGATTAATGCTCTTGGCTGGGTGTGCT 66923
QY 667 CAAGCACTCGGCATTGACGAGAAACAGCAATTTAAGGAATGCTGAATGCGCCGACAT 726
Db 66922 CAAGCACTCGGCATTGACGAGAAACAGCAATTTAAGGAATGCTGAATGCGCCGCTGAT 66863

QY 727 CCGGAGCAATGAGAAATCTTCGCTGATCATGCCAGCGCTGGGCACTTGTATAT 786
Db 66862 CCGGAGCAATGAGAAATCTTCGCTGATCATGCCAGCAAGAAATCCCGGACATTTGCTCAAC 66803
QY 787 GGGTTTCCCGAAACGACGCTTCTTCACTTTGATATATGAAACGCTGTAAAAAGAAATC 846
Db 66802 GGGTTTCCCGCAATGACGAGCTTCCACTTTTAAACTTTGGAAGCCTGTAAAAAGAAATA 66743
QY 847 GGTACCAGCCAGTATCCGATCATCATGTAACGCGCGAGACGCTGTGATCGG 906
Db 66742 GGTATCTTACGATCAGCCGATGCTTATTAATGAACTCCGCGCGACAGGTTGACAGA 66683
QY 907 ACAACGCAATTCGCAATGACGTAATTCCTTAATTAAGCAAGTAAGTATCTTAATC 966
Db 66682 ACAACGCAATTCGCAATGACGTAATTCCTTAATTAAGCAAGTAAGTATCTTAATC 66623
QY 967 GGTGAACCAACAGAACCGATGCTTAAAGCCTTAAGAGAAAGCAAAATTCCTGCAGACAA 1026
Db 66622 GGAAGAACCAACAGAACCGATGCTTAAAGCCTTAAGAGAAAGCAAAATTCCTGCAGACAG 66563
QY 1027 CTGATGACCTAGATTAAGTCAACAGATGAATTAAGAAATTTGTAAGAAAGAAATG 1086
Db 66562 CTGTTGATTTTGAACCAATTAACGAGAAAGAAATTAAGTTCATGCTGAATAACAGCTT 66503
QY 1087 CACAACTGTATATATGAGCTGCGCAATATTCATGTCGCGAGACCTTTAATTGAA 1146
Db 66502 GAGGCGCGCTTATTAAGGAGTGGAAATATTCAGAGACAGCGAGCGCTTCATTGAA 66443
QY 1147 AAAATCAAGATTCAGATTAAGCAAGCTGTTAGC 1182
Db 66442 AAAATCAAGATTCAGATTAAGCAAGCTGTTAGC 66407

RESULT 10
CP000002_36/c
WPCOMMENT

Sequence split into 43 fragments LOCUS CP000002 Accession CP000002

Fragment Name	Begin	End
CP000002_00	1	110000
CP000002_01	100001	210000
CP000002_02	200001	310000
CP000002_03	300001	410000
CP000002_04	400001	510000
CP000002_05	500001	610000
CP000002_06	600001	710000
CP000002_07	700001	810000
CP000002_08	800001	910000
CP000002_09	900001	1010000
CP000002_10	1000001	1110000
CP000002_11	1100001	1210000
CP000002_12	1200001	1310000
CP000002_13	1300001	1410000
CP000002_14	1400001	1510000
CP000002_15	1500001	1610000
CP000002_16	1600001	1710000
CP000002_17	1700001	1810000
CP000002_18	1800001	1910000
CP000002_19	1900001	2010000
CP000002_20	2000001	2110000
CP000002_21	2100001	2210000
CP000002_22	2200001	2310000
CP000002_23	2300001	2410000
CP000002_24	2400001	2510000
CP000002_25	2500001	2610000
CP000002_26	2600001	2710000
CP000002_27	2700001	2810000
CP000002_28	2800001	2910000
CP000002_29	2900001	3010000
CP000002_30	3000001	3110000
CP000002_31	3100001	3210000
CP000002_32	3200001	3310000
CP000002_33	3300001	3410000

/db_xref="GI:1894755"
/translation="MDQEAAYSEYKARQTTADQENVPSCKEIIIAVRNEDNIIAY
KTNIGREDUPTALSEAKSGKLAVDV"
/complement(221..227)
/gene="Yw8A"
/complement(334..921)
/gene="Yw8B"
/complement(334..369)
/gene="Yw8B"
/complement(371..907)
/gene="Yw8B"
/codon_start=1
/transl_table=1
/product="unknown"
/protein_id="CAB07460.1"
/db_xref="GI:1894756"
/db_xref="InterPro:IPR001452"
/db_xref="InterPro:IPR003646"
/db_xref="UniProt/TREMBL:P96729"
/translation="MNKPTKLFSTLALAAGMTAAAGAGTTHAQOPEPTVSINDIYS
YPIISYLSAEALAVRTKPSASQKADPTLHGDLSKLSFENADMAYKYKNGCTGPV
STHYIVKAATVTKTKTKVTSADGKSKTLPADTVSYFLGMSKTNKGFPDWVFV
DYGGTGYMKTGDLHMTK"
/complement(914..921)
/gene="Yw8B"
/complement(join(989..1906,1912..1919))
/gene="rb8B"
/complement(989..1906)
/gene="rb8B"
/note="part of the ribose transport operon"
/citation=[1]
/codon_start=1
/transl_table=1
/product="periplasmic substrate-binding protein"
/protein_id="CAB07461.1"
/db_xref="GI:1894757"
/db_xref="GOA:P36949"
/db_xref="InterPro:IPR001761"
/db_xref="UniProt/Swiss-Prot:P36949"
/translation="MKKAVSVILTLSTFLTLACSLPEPQWAKPSNSGNKEFTGLSY
STLNPPFVSLKGIKEKAKRGKMKVIIVDQNDSSKQTSDEDLIQGVNALLINPT
DSSAISTAYSEBANAVPVVTIDRSAGEKVTETVASDNRVKGEMAAFIADKLKGKA
KVALEEGVSGASATREBGSFPHNIADQGLVTKQSADPDRTKGLTWENTLQGHPI
QAVRHNDEMALGALAINSSKDIIVTGPQNMDALASIDRKLSAIVAAQPELICK
LATRAADDILHGKVKQTTISAPLKLRTQK"
/complement(1559
/citation=[1]
/replace="g"
1820..1838
/citation=[1]
/replace="tcgttcgccatcgga"
/complement(1912..1919)
/gene="rb8B"
/complement(1918..2886)
/gene="rb8C"
/complement(1918..2886)
/note="part of the ribose transport operon"
/citation=[1]
/codon_start=1
/transl_table=1
/product="membrane transport protein"
/protein_id="CAB07462.1"
/db_xref="GI:1894758"
/db_xref="GOA:P36948"
/db_xref="InterPro:IPR001851"
/db_xref="UniProt/Swiss-Prot:P36948"
/translation="MKTEQLQTEQKRIHFDVWQKLGFLPLIVITVSLINSPLE
PLNLTNLRQVAINGLIFGMPVILITGIDLSVGAIIALSALVAGMTVSGVDPVLA
IILCCIIGAVLGINGLITKGMKAPATATMTVPFGLILVYTDGPIITGDTNG
FQMGGRGFLGIPVAITMVALFVILWLLHKTFFGRTVAIIGNEKALISGIKVTR
VKMIVISLAGLLSLAGAILTSRLHSAQPTAGSEYELDAIAAVVLGTSLSGGRRIY

GTLLIGLLIQTNNGLNLGVSSFYQLVWGVILLAVLLDRKKA"
2065..2066
/citation=[1]
/replace="cg"
2117..2118
/citation=[1]
/replace="cg"
2188..2389
/citation=[1]
/replace="cgagc"
2485..2486
/citation=[1]
/replace="gc"
2846
/citation=[1]
/replace="c"
/complement(join(2888..4369,4374..4383))
/gene="rbsA"
/complement(2888..4369)
/gene="rbsA"
/note="part of the ribose transport operon"
/citation=[1]
/codon_start=1
/transl_table=1
/product="ATP-binding transport protein"
/protein_id="CAB07463.1"
/db_xref="GI:1894759"
/db_xref="GOA:P36947"
/db_xref="InterPro:IPR003439"
/db_xref="InterPro:IPR003593"
/db_xref="UniProt/Swiss-Prot:P36947"
/translation="MQIEMKDTHKTFGKNQVLGVSFQLPMPGEVHALMGNGAKSTL
MNLTGHRKADKGQISINGEYFENSPKEAEOGIARFHOELNIMPETVLENLFCK
EISKLGVLQTRKALAKEOPDKLSVSLDDEAGCSVGQOOMEIAKALMTNARY
IIMDEPRAALTERISGLFEVITLAKNGSIVYISRMETIFRICRITIMRDQKY
DTNISTEDDEVAKRWAGRELTERYPKTPSGDKVEVKNASVKGSFEDVSFYVS
GEIVGSGLMAGRTKEMMALFGVDRLDGEIWIAGKKTAKPOEVKKGGLFTEN
RKDEGLLDITSRENIALPNLSFSFPGILDHREAEFVLLIKRLTIKTASPTTHAR
HLSGNOOKVIAKVIIGIPKVLIDPEKGVDPVGAKREIYTLNNELTBERGVALIMVS
SELPEITGMSDRITVHEGRISGEIHRAREATQERIMTLATGR"
/complement(2895..2899)
/gene="rb8C"
/complement(2895..2899)
/gene="rb8C"
/complement(2895..2899)
/citation=[1]
/replace="ttcttcgccgttcg"
3220..3343
/citation=[1]
/replace="tc"
3397..3399
/citation=[1]
/replace="ta"
4241..4245
/citation=[1]
/replace="ccgtg"
/complement(4374..4383)
/gene="rbsA"
/complement(4385..4780)
/gene="rbsD"
/complement(4385..4780)
/gene="rbsD"
/note="part of the ribose transport operon"
/citation=[1]
/codon_start=1
/transl_table=1
/product="membrane transport protein"
/protein_id="CAB07464.1"
/db_xref="GI:1894760"
/db_xref="GOA:P36946"
/db_xref="InterPro:IPR007721"
/db_xref="UniProt/Swiss-Prot:P36946"
/translation="MKKGIIINSHLAKLIDLGHITDKIIVADAGLPVPDGVLLKIDLSL

gene
 CDS
 KPGIAPRODTAALAEEMVAKVIAAEIKASNOENKFLLENFSEQIEVLSHEEFK
 LITIDAKAVIRIGETPIYANCILQAGVLP"
 complement (4777. .5658)
 /gene="fbak"
 complement (4777. .5658)
 /gene="fbak"

Query Match 62.0%; Score 733.2; DB 1; Length 8164;
 Best Local Similarity 99.6%; Pred. No. 1.2e-152;
 Matches 735; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 7 TGGTACTCATTATATGCGCTGCTGTCTGTCATCTGTCATCGGAAATATTGAAAAACGACGA 66
 DB 7427 TGGTACTCATTATATGCGCTGCTGTCTGTCATCTGTCATCGGAAATATTGAAAAACGACGA 7486
 QY 67 CATCAGAAAAACATTGATGCCCTCCCTGTTGGGTGAATATTAAACGGCATCCGCGAAAA 126
 DB 7487 CATCAGAAAAACATTGATGCCCTCCCTGTTGGGTGAATATTAAACGGCATCCGCGAAAA 7546
 QY 127 TCGACTGTGACAAAGCTGACAAACCGGAATATTAAAGAGCCGTTACAAAGCTGTGGA 186
 DB 7547 TCGACTGTGACAAAGCTGACAAACCGGAATATTAAAGAGCCGTTACAAAGCTGTGGA 7606
 QY 187 AAAACAAACGAAACAGATGCAAGAAATGATTACTGGGACACACCGGAGGAAAAAGCGAAT 246
 DB 7607 AAAACAAACGAAACAGATGCAAGAAATGATTACTGGGACACACCGGAGGAAAAAGCGAAT 7666
 QY 247 AAACGGAACCTCAGAGGCGCAATATCGAGAGCAAAAAAGAGTATGAGAGAAACAGTA 306
 DB 7667 AAACGGAACCTCAGAGGCGCAATATCGAGAGCAAAAAAGAGTATGAGAGAAACAGTA 7726
 QY 307 GAAAGAGGGGCTAACGCGATTGTCAATGATGATGATGATGATGATGATGATGATGATGATG 366
 DB 7727 GAAAGAGGGGCTAACGCGATTGTCAATGATGATGATGATGATGATGATGATGATGATG 7786
 QY 367 ATCTTTACAGAGAACTTTGACAGGCAATATCGCGCTGATGATGATGATGATGATGATGATG 426
 DB 7787 ATCTTTACAGAGAACTTTGACAGGCAATATCGCGCTGATGATGATGATGATGATGATGATG 7846
 QY 427 CATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 486
 DB 7847 CATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 7906
 QY 487 CCTTATATGCGCATCTTGTCTATTAACAGATGATGATGATGATGATGATGATGATGATGATG 546
 DB 7907 CCTTATATGCGCATCTTGTCTATTAACAGATGATGATGATGATGATGATGATGATGATGATG 7966
 QY 547 GCAAAAAGCAAAACCAAAAGTCAATCTGATGATGATGATGATGATGATGATGATGATGATG 606
 DB 7967 GCAAAAAGCAAAACCAAAAGTCAATCTGATGATGATGATGATGATGATGATGATGATGATG 8026
 QY 607 TTACGTAATTTTGAATACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 666
 DB 8027 TTACGTAATTTTGAATACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8086
 QY 667 CAAGCACTGGCATTTGACAGAAACAGCATTTTAAAGGAATGCTGAATGCGCGCAGAT 726
 DB 8087 CAAGCACTGGCATTTGACAGAAACAGCATTTTAAAGGAATGCTGAATGCGCGCAGAT 8146
 QY 727 CCGGAGCAATGAAATT 744
 DB 8147 CCGGAGCAATGAAATT 8164

RESULT 12
 AX434279 819 bp DNA linear PAT 28-JUN-2002
 LOCUS AX434279
 DEFINITION Sequence 2694 from Patent WO0229113.
 ACCESSION AX434279
 VERSION AX434279.1 GI:21659087
 KEYWORDS
 SOURCE Bacillus licheniformis
 ORGANISM Bacillus licheniformis

REFERENCE
 1 Berka, R. and Clausen, I. G.
 METHODS FOR MONITORING MULTIPLE GENE EXPRESSION
 TITLE Patent: WO 0229113-A 2694 11-APR-2002;
 JOURNAL Novozymes Biotech, Inc. (US) ; Novozymes A/S (DK)
 FEATURES
 SOURCE
 1. 819
 /organism="Bacillus licheniformis"
 /mol_type="unassigned DNA"
 /db_xref="taxon:1402"

ORIGIN
 Query Match 46.4%; Score 548.4; DB 6; Length 819;
 Best Local Similarity 79.9%; Pred. No. 2.3e-111;
 Matches 645; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

QY 24 CTGTGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 83
 DB 4 CTGTGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 63
 QY 84 TGCCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 143
 DB 64 TGCCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 123
 QY 144 GACACCGGCAATTTAATGAAAGCGGTACAGACTGTGTAAGAAACAAACAGAAACAGA 203
 DB 124 AACCAACAGGCAATTTAATGAAAGCGGTACAGACTGTGTAAGAAACAAACAGAAACAGA 183
 QY 204 TGCAGATGATTTTAACTGCGACACACCGGAGAAAAAGCGGATTAACGGAACCTCAGGG 263
 DB 184 CGCAAGATGATTTTAACTGCGACACACCGGAGAAAAAGCGGATTAACGGAACCTCAGGG 243
 QY 264 GCCGAATATCGAGAGCAAAAGAAAGTATGAGAGAAACAGTAAAGAAAGGGGCTTAAGC 323
 DB 244 GCCGAATATCGAGAGCAAAAGAAAGTATGAGAGAAACAGTAAAGAAAGGGGCTTAAGC 303
 QY 324 GATTTGCTGATTAATGCAATGCTGTTAACCGGATTTATCAATCTTTCAGAGAAACT 383
 DB 304 GATTTGCTGATTAATGCAATGCTGTTAACCGGATTTATCAATCTTTCAGAGAAACT 363
 QY 384 TCTGAGGCAATATCGGCGTATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATG 443
 DB 364 GCTTACGCTAATATCGGCGTATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATG 423
 QY 444 GCCGACGCTTGAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATG 503
 DB 424 ACCGACTTTGATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATG 483
 QY 504 TGTCTATCAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATG 563
 DB 484 GGTATTTCTGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATG 543
 QY 564 AAAAGTCATCATGCTGATTAATCAAAATTAACAGATGATTTTACGTAATTTTGAATA 623
 DB 544 AAAAGTCATCATGCTGATTAATCAAAATTAACAGATGATTTTACGTAATTTTGAATA 603
 QY 624 CATGTAATTCCTGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATG 683
 DB 604 CATGTAATTCCTGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATG 663
 QY 684 CGAAGAAACAGATTTTAAAGGAATGCTGAATGCGCGCAGATCCGGAGCAATGAAAT 743
 DB 664 CGAAGAAACAGATTTTAAAGGAATGCTGAATGCGCGCAGATCCGGAGCAATGAAAT 723
 QY 744 TCTTCGCTGATCAGTCGAGGAGCGCTGGGCACTTTGTAATGAGTTTGGCCGAAACGA 803
 DB 724 TCTTCGCTGATCAGTCGAGGAGCGCTGGGCACTTTGTAATGAGTTTGGCCGAAACGA 783
 QY 804 CGCTTCTTCTACTTGAATATGAA 830
 DB 784 CGAGCTTCACCTTTAAACATTTGAA 810

/db_xref="GI:20520288"
/translation="MVKLGKKKQOKRTEGYDLDFISRVPOGGIKENEKXIAOGDCY
VACHVYSLADIPIPLMTALMNNKNDITSSVDPAITANKEEVVKDINSITLKRDMTS
ERRSTDRDADAMELONLTPARSTQGEIYKLYKLRVYIYDPLBLOEKIGDIXKE
IAGNYKAOVYFKQKEBQTLFASYSDDQIYLVGKSYPLPSKNIQGLGIFPHNDLK
DPRIVYIGQSTGAFILDPFSTGRTSFSGF1GKGAKAKSTLLKQLEGVLAQDC
FIRBFDNARDVYTVVOOGGKIIDLQVQDEBEVETGINLEVFATKITVYGAUDEK
GSFLOHISKVTNMIPLNTEPDDTVIOEFRGLIYAFYEMGLVPOKSGDRPYKTYGP
PNHYIILEDPEYKISNLKLEAGATPQRHDELAIOVEDMITVYGDWFGNHTLLKN
ENBOIVPEDIQISKYKSPVNCOLFATLTITMSHALKNGBOMYLAEREKLSIEDVK
YPMVFLDECHAVINSQNEFANRYVLEFEREKRPAGIFERTQSPQENLPFGSTVDY
STYMAIFELTQYKIFLNNDNSVLETLKSLBLSLSEFRLPELKREALIVQSVSTE
TYNVPFDEKQYLFKFGQ"

gene
CDS
complement (6684 . 7358)
/gene="BXB0009"
complement (6684 . 7358)
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein, (pX02-10) "
/protein_id="AAM26170.1"
/db_xref="GI:20520289"
/translation="MKMKVATAEAKRVNRKQISEKNGKOSFISTIPASFKNXTDAS
IVDIAPYRLLEQKTGILVDHNNILQVLYKTTDLISNODDLKRFNMQLTSLGVYH
EPFLLSLTSETTEEOQVYKWRMALARFGMSQSEVSKKEEHLAYORXSLAINLR
VLVREKNLKEALFIYVYGNKETELIKVVKMKRIGRQFNLQNMKAKEVEKLIFFKIQ
NMSEB"

gene
CDS
complement (7416 . 7730)
/gene="BXB0010"
complement (7416 . 7730)
/gene="BXB0010"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein, (pX02-11) "
/protein_id="AAM26171.1"
/db_xref="GI:20520290"
/translation="MEMMRNPKNTKQEIKLAFYIITDAITAIMLVLASYMPKVPVVG
GGRIMFYVLFQSTGRLFLCIKPHNSPTNRNIFVILDLMKMNKXHPLEVNITISSEK
RK"

gene
CDS
complement (7733 . 7936)
/gene="BXB0011"
complement (7733 . 7936)
/gene="BXB0011"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein, (pX02-12) "
/protein_id="AAM26172.1"
/db_xref="GI:20520291"
/translation="NNNNIPALIMGVAIVUSIATPFKQIFTIKRRKKKEPWVWTV
TNAEYLDREDFWKGEKQDKKEHD"

gene
CDS
complement (7938 . 8192)
/gene="BXB0012"
complement (7938 . 8192)
/gene="BXB0012"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein, (pX02-13) "
/protein_id="AAM26173.1"

Query Match 42.9%; Score 507.2; DB 1; Length 94829;
Best Local Similarity 64.8%; Pred. No. 1.6e-102;
Matches 752; Conservative 0; Mismatches 408; Indels 0; Gaps 0;

QY 1 ATGGGCTGTTACTATATAGCCGTGTGTCATACGTCATGCGAATATTAGAAAA 60
DB 56792 ATGATCTTCATATAGCTATAGTACAGTGTTTGAATTATTATGCTATATGGAACAA 56793
QY 61 CGACGACATCAGAAAAACATTGATGCCCTGTTGGGGTGAATATTAAACGACATCCGC 120

DB 56732 CGTTCACATCAGAAAAGGCTCAATTCTATCCCAATTCGAGTAAACATAAATGCAATTGCA 56673
QY 121 GGAATAATGCACTGTGACAAAGGCTGACAAACGGAATATTAATAGACCGGTTACAAGACT 180
DB 56672 GGTAAATCTACCGTTACAAAGCTAAATTAACAGTGCTTGTACAAAGAACGAATATTAAGACT 56613
QY 181 GTTGGAAAAACACAGAACAGATGCAAGAAATGATTTACTGGGACACACCGAGGAAAAAG 240
DB 56612 GTAGGAAAAACAATGCTGATCTGCCCGAATGATATATATGTTTACTGACAGAGACCAA 56553
QY 241 CCGATTTAAACGGAAACCTCAGGGGCGGAATATGGAAGAGAAAAAGAAAGTCATGAGAA 300
DB 56552 CCGATTTAAGCGCGTAAAGAGGCTCTTAATATGCTGAGCAACGAGGTAATGTTAAAGAG 56493
QY 301 ACACTAGAAAGAGGGGCTAAACGCGATTTGACATGCAATGCTGTTTAAACCGAGATTAT 360
DB 56492 GCTGCTGATTTTGAAGACGAAGACATTAATTTGTAATGTAATGGACATTCACCGGATTAT 56433
QY 361 CAAATCAATCTTTTCAGAAACCTTCTGACGGCCAAATATCGGCGTCATTGTGAATGTTTTA 420
DB 56432 CAAATTAATCTTCCAAATTAATAATGATTCAGCAAAATGTTGAGATGTTAAATGTTTTA 56373
QY 421 GAAGACATATGATGATGTCATGAGGCGCCAGCGCTGATGAAATTTGAGAAAGCTTAAACCGCT 480
DB 56372 GAAGATCATATGATGATGTTATGAGCACTAACCTTGACAAAGATGCTGAAGCTTTCACGTCT 56313
QY 481 ACAATTCCTTAATATGCGCATCTTGTCATTACAGATAGTGAATATPACCGAGTTCTTTAA 540
DB 56312 ACCATTCATATATGACATTTTATGTCATATGTAAGAAATGTAATCTTGATTAATTAA 56253
QY 541 CAAAAGCAAAAGAACAAACAAACAAAGTCATCTGATTAATCTGAAATTAACAGAT 600
DB 56252 GAGGTGAGAAAGAGAAATACAAAGATGATTTGTCGATATTTAGAAATTTACAGAA 56193
QY 601 GAGTATTTAGCTATTTTGAATATACATGTAATTCCTGTATTAACGTTCTCTGCGCGCTGGT 660
DB 56192 GAAATCTTACAAAAATTTGATTAATGATGATGCTTCCCAATATATGATGCTTGCTTTAGCG 56133
QY 661 GTGGCTCAAGCACTCGGCATTTGACGAAGAAACAGACATTTAAGGAAATGCTGAATGCGCG 720
DB 56132 GTAGCAGAGGCTCTTGAGATTGATGAGAAACAGACATTCGCGGTATTTGTAATGTCAT 56073
QY 721 CCAATCCGAGGAGCAATGAGAAATCTTCCGCTGATCAGTCGAGCGAGCCTGGCACTTT 780
DB 56072 CCGATCCAGAGAGCAATGAGAAATTAACAGTTTCTGACCAATTAAGCCTGCGCTTCTTC 56013
QY 781 GTTAATGGGTTGCGCGCAACAGACGCTTCTTCTATCTTTGAATATATGAAACCGTGAATA 840
DB 56012 GTTAATGGTTTTGCAGGGAATGATCCCTCAACATTAACGTATTTGGGAACGTTGTGAT 55953
QY 841 GAAATCGGTTACCGGACCGATGATCCGATCATCATCATGAACCTCCGCGACGACCGTGT 900
DB 55952 GATTTTGATATATGTAATCTAGCTCCAAATTTGTAATTTAGAAATTCGCCCGCTGACCGGTT 55893
QY 901 GATCGGACACAGCAATTCGCAAAATGACGTATTTGCTTAATTTGAAGCAAGTGAACGTATC 960
DB 55892 GATGTACTAGCAAGTTTGCTAGGAGATGTTTGCATATATTAAGCGGAAATAGTATAT 55833
QY 961 TTAATCGGTGAACACAGAACCAATCGTAAGACCTATGAAGAAAGCAAAATTCCTGCA 1020
DB 55832 GCGATTTGAGAAACGACCTGACCTTAATCAAGTCTTTTGAAGAAAGAAATTTCCAACG 55773
QY 1021 GACAAACGATGACCTAGAGTATATAGTCAACAGATGAATATGAAATGTTTAAAGAA 1080
DB 55772 CAAGGTAATGGAATTTGAAGGCTGGTCAACAGTGAATTAATGTTCTCGATCGTCCA 55713
QY 1081 AGAATGACAAACCGTGCATATATATGAGCTGCGCAATATTCATGTTGCGCGACAGACCTTTA 1140
DB 55712 TATTTAAAAAATCGGATGTAATGAGATGAGTGAATATATATGTTGACGCTGACCATTA 55653
QY 1141 ATTGAAAAATTCACGAATA 1160
DB 55652 ATCGATATGATTAATGAAGA 55633

RESULT 15	AE017335	94830 bp	DNA	circular BCT 09-JUL-2004
LOCUS	AE017335/c			
DEFINITION	Bacillus anthracis str. 'Ames Ancestor'	plasmid pXO2, complete		
ACCESSION	AE017335			
VERSION	AE017335.3	GI:50118566		
KEYWORDS				
SOURCE				
ORGANISM	Bacillus anthracis str. 'Ames Ancestor'			
REFERENCE	Bacillus anthracis str. 'Ames Ancestor'			
AUTHORS	Bacillus anthracis str. 'Ames Ancestor'			
TITLE	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus; Bacillus cereus group.			
JOURNAL	1 (bases 1 to 94830)			
REFERENCE	Ravel, J., Rasko, D.A., Shumway, M.F., Jiang, L., Cer, R.Z.,			
AUTHORS	Federova, N.B., Wilson, M., Stanley, S., Decker, S., Read, T.D.,			
TITLE	Salzberg, S. and Fraser, C.M.			
JOURNAL	Bacillus anthracis comparative genomics			
REFERENCE	2 (bases 1 to 94830)			
AUTHORS	Ravel, J., Rasko, D.A., Shumway, M.F., Jiang, L., Cer, R.Z.,			
TITLE	Federova, N.B., Salzberg, S. and Fraser, C.M.			
JOURNAL	Direct Submission			
REFERENCE	Submitted (17-MAY-2004) Microbial Genomics, The Institute for			
AUTHORS	Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850,			
TITLE	USA			
JOURNAL	Sequence update by submitter			
REFERENCE	On Jul 9, 2004 this sequence version replaced gi:47552342.			
AUTHORS	Location/Qualifiers			
TITLE	1. 94830			
JOURNAL	/organism="Bacillus anthracis str. 'Ames Ancestor'"			
REFERENCE	/mol_type="genomic DNA"			
AUTHORS	/strain="Ames Ancestor; A2084"			
TITLE	/db_xref="taxon:261594"			
JOURNAL	/plasmid="pXO2"			
REFERENCE	/note="MLVA types GT62 and A3.b"			
AUTHORS	/locus_tag="GBAA_PXO2_0001"			
TITLE	complement(543..1139)			
JOURNAL	/locus_tag="GBAA_PXO2_0001"			
REFERENCE	/note="identified by Glimmer2; putative"			
AUTHORS	/codon_start=1			
TITLE	/transl_table=11			
JOURNAL	/product="hypothetical protein, (pXO2-03)"			
REFERENCE	/protein_id="AAI28931.2"			
AUTHORS	/db_xref="GI:47552343"			
TITLE	/translation="MTYVKLFGFRGQIHVDVDFQROYPEYPRADVLRAIVLID			
JOURNAL	EKKREKGGDDVSLKKEVDFDKOSVNTMKQMDVLIKUNMELIAQSGNSLSEKLEQ			
REFERENCE	EYVSEIASSVTYKSEKGFPTVYREPIEENVOEHRIQREVKTSFVKPLPMITDDV			
AUTHORS	QIVMKKGKPHOVAPRTLGTYGQIEIWEQIPEHRKRLKLE"			
TITLE	complement(1151..1438)			
JOURNAL	/locus_tag="GBAA_PXO2_0002"			
REFERENCE	complement(1151..1438)			
AUTHORS	/locus_tag="GBAA_PXO2_0002"			
TITLE	/note="identified by Glimmer2; putative"			
JOURNAL	/codon_start=1			
REFERENCE	/transl_table=11			
AUTHORS	/product="hypothetical protein, (pXO2-04)"			
TITLE	/protein_id="AAI28932.2"			
JOURNAL	/db_xref="GI:47552344"			
REFERENCE	/translation="MEIHVRNVDPYHLKEIDRCKEIGKKLGRYYRWEXYINMMFQGH			
AUTHORS	FNOEYRNKEDKDEAVSNVITLDROSDLGIVADITHLVAAIMKLNEE"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..1768)			
JOURNAL	/locus_tag="GBAA_PXO2_0003"			
REFERENCE	complement(1652..1768)			
AUTHORS	/locus_tag="GBAA_PXO2_0003"			
TITLE	complement(1652..17			

```
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein, (pX02-09)"
/protein_id="AAT28938.2"
/db_xref="GI:47552350"
/translation="MVKLKKEQKQKRTGYDLDFISHPQPGGIKFNEKYIAODCY
VACHVYSIADIPMLTALMNNKDTISVDVATANKEEVKNIDNITELKDRMTS
ERSTDDPAHMELONTDPASITQOGEIKVLRITVYDPULEOLERIGDIKKE
IAGONYKAOVYVFKQEKEMQTLFASVDQIEILYKSGYPIPSKRIIGFIFPHQDLK
DPRGIYLGQSTGAFILDPFSTGTRTSFGSFIKGNAGKSTLLKOLEGLAVKDC
FIRGDKARADYVTVQOQGGKIIDLGVQDSBEVETGMINPLEVATKITVTGAVDEK
GSFIQISKVTNMIREFDDIVIOFRKHLVAFYEMGLVPOQSGDRYKATGYD
ENHVIIEEDFYKYSNLKLEAGATPQRHDLALIOVEDMITVYGDMDGHNTLKNF
ENHOVPEPDIGISKYDQSVNCOLEFALTIIMSHALKNGBOMKYLREKNSLIEDYK
YPMVFLDECHVINSQNEFANRYVLEFERNRKKFPAGIFPTQSOEHLPPGSDVDY
STMKIFELTQKIFLNMDSVLETLKVLDSLESEFRILPELKREALIVQVSTSE
TYNMFDPDEKQLERFKGGQ"
/complement(6684..7358)
/locus_tag="GBAA_PX02_0009"
/complement(6684..7358)
/locus_tag="GBAA_PX02_0009"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein, (pX02-10)"
/protein_id="AAT28939.2"
/db_xref="GI:47552351"
/translation="MKMKVATAEAKKNRKOIESEKNGKOSFISTIPASFKNITDAS
IVDIAPYRLBQKTGILVDHRNNLQVYIKVITDLSNMODLKFPMQOLSLCRVYH
EPFKILSTYETEQQVYWKEMALRFQGMRSQVSEKKEHLWYQKYSIALINLR
VLAVKKNLKEALFIVVYQKNETELIKVKNMKRYGQFNLQNMKAKEVERKLFLKIQ
NMNSEM"
/complement(7416..7730)
/locus_tag="GBAA_PX02_0010"
/complement(7416..7730)
/locus_tag="GBAA_PX02_0010"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein, (pX02-11)"
/protein_id="AAT28940.2"
/db_xref="GI:47552352"
/translation="MEMMRNPKNTKQIKIAFFYIIDAIIALMLVLSYMPKVPVQ
GGRIMFYVLRGTGFLPICIKPHNSPTNRNIFVILDLKMNKNVHPLEVNTISSEK
RK"
/complement(7733..7936)
/locus_tag="GBAA_PX02_0011"
/complement(7733..7936)
/locus_tag="GBAA_PX02_0011"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein, (pX02-12)"
/protein_id="AAT28941.2"
/db_xref="GI:47552353"
/translation="MNNWIFALIMGLVAIVLSIITPFKQITFIKRRKKKEPVWTV
INAEYLRDPFKGKEQKDKEND"
Query Match 42.9%; Score 507.2; DB 1; Length 94830;
Best Local Similarity 64.8%; Pred. No. 1.6e-102;
Matches 752; Conservative 0; Mismatches 408; Indels 0; Gaps 0;
QY 1 ATGGGCTGGTTACTATTAAGCTGTGCTGTCATACGTCATCGAATATTAGAAAA 60
Db 56792 ATGATCTTCATATAGTATATGTAAGTGTGTTGATTATTATGTAATGGAACA 56793
QY 61 CGAGCAATCAGAAAAAATTGATGCCCTCGCTGCGGTAATATTAAAGCAATCGC 120
Db 56732 GGTTCCTACAGAAAAGGCTCAATTCATCCCAATTCGAGTAAACATTAATGAAATTCGA 56733
QY 121 GGAATAATGCACTGTGACAAAGCTGACAAACCGAATATTATAGAACCGGTTACAGACT 180
Db 121 GGAATAATGCACTGTGACAAAGCTGACAAACCGAATATTATAGAACCGGTTACAGACT 180
```

```
Db 56672 GGTAAATCTACCGTTACAGACTAATTATGAGGTGTTGACAGAAAGCAAAATATTAAGACT 56613
QY 181 GTTGGAACAAACAGGAACAGATGCAAGATGTTACTGGGACACACCGGAGGAAAA 240
Db 56612 GTAGGAAAAACACTGTGACCTGCGCGAATGATATATGTTTACTTGACAGAGACAA 56553
QY 241 CCGATTAAACGGAAACCTCAGGGGCCGAATATCGAGAGCAAAAAGAGTCATGAGAAA 300
Db 56552 CCGATTAAAGCCGCTAAAGAAAGTCCTTAATATGCGTAGCAACGAGGTAAGTAAAGAG 56493
QY 301 ACAGTAGAAAGGGGCTAACCGCATTTGATGTAATGCAATGCTGTTTAAACCAAGATTAT 360
Db 56492 GCTCCTGATTTAGAGCAAGAGCACTTAATTTGTAATGTAATGCAAGTTCAACCGAATAT 56433
QY 361 CAATTCATCTTTGAGGAAGAACTTCGAGGCCAATATCGGCGCATTTGTAATGTTTAA 420
Db 56432 CAATTCATCTTTGAGGAAGAACTTCGAGGCCAATATCGGCGCATTTGTAATGTTTAA 56373
QY 421 GAGACCATATGATGATGTCATGGGGCCGACGCTTGATGAAATTCGAGAGCGTTTACCGCT 480
Db 56372 GAGATCATATGATGATGTTATGGGACCTTACACTTGACGAAAGTGAAGCTTTCACCTGCT 56313
QY 481 ACAATTCCTTAATATGCGCATCTTGTCATTAACAGATAGTGAATATACCGAATCTTTTAA 540
Db 56312 ACCAATTCCTTAATATGACATTATGTCATTAAGAAAGTGAATCTTGGAATTTTAA 56253
QY 541 CAAAAGCAAAAGAACGAAACACAAAGTCATCTGCTGATACCTGAAATATACAGAT 600
Db 56252 GAGGTGACAGAGAGAGAAATACAAAGATGATGTTGCGGATTAATCTTGAATTTACGAA 56193
QY 601 GAGTATTTTACCTAATTTTGAATATACATGATATCCCTGATTAACGCTTCTGCGCGTGGT 660
Db 56192 GAATTTTACGAAATTTGATATACATGCTGCTCCAGATATATGATGATGCTTGTAGCG 56133
QY 661 GTGCTCAGACACTCGGCATGACGAAAGAACAGCAATTTAAAGGAATGCTGAATGCCGCG 720
Db 56132 GTACACAGAGGCTCTTGCGATGATGATGAGGAAACAGCAATTCGCTGATGTAAGTGCAT 56073
QY 721 CCAATCCGGAGAGAAATGAGAAATCTTCGCTGATCAGTCCGAGCGAGCCGCGACCTTT 780
Db 56072 CCGAATCCAGAGAGAAATGAGAAATTAACGTTTCTGACCAATGTAAGCCGCGCTTTC 56013
QY 781 GTTAATGCGTTTGGCGGAAACGACGCTTCTTCTACTTTGAATATATGAAACGTGTAAA 840
Db 56012 GTTAATGCGTTTGGCGGAAAGATCCCTCATCAATTAAGTATTTGGAACGATGCGAT 55953
QY 841 GAAATCGGTTACCCGACCGATGATCCGATCATCATGAACTGCCGCGACACCGCTGTC 900
Db 55952 GATTTTGGATATGATTAATCTAGCTCCAATGTAATTTATGAAATTCGCCCGCTGACCGGCT 55893
QY 901 GATCGGACACGCAATTCGAAATGAGGTATGCTTATATGTAAGCAAGTGAACGTGATC 960
Db 55892 GATCGTACGACGAGTTTGTGAGGAGTGTGTCATATATTTAAAGCGAAATATGTTAT 55833
QY 961 TTAATCGGTGAAACAGAAACGATGCTAAACCTATGAGAAAGCAAAATTCCTGCA 1020
Db 55832 GCGATGAGAAAGACATGCGACCTATTAACAAGTCTTTGAAAAGAGATATTCACAAG 55773
QY 1021 GACAAATCGATGACCTAGAGTATAGTCAACAGATGAATATATGAAATTTGTAAGAAA 1080
Db 55772 CAAGAGTATGGAACCTTAGAGGCTGCTCAACAAGTAAATATATGTCCTGATCCGTCA 55713
QY 1081 AGAATGCAACCGTGTCAATATAGCGGTGGGCAATTTATCATGGTCCGCAAGACCTTTA 1140
Db 55712 TATTTAAAAAATCGAATGTAATGAGGTGGTAAATTCATGGTGAAGCTGAGCCATTA 55653
QY 1141 ATTGAAAAATCCAGAAATA 1160
Db 55652 ATCGATATGATTAAGAAAG 55633
```

Search completed: February 27, 2006, 11:11:01
Job time : 6811.28 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:36:56 ; Search time 6671.32 Seconds
(without alignments)
8289.569 Million cell updates/sec

Title: US-10-789-164-1

Perfect score: 1182
Sequence: 1 atgggcgtgcttaccatcatat.....aggaagcagctcgtgaac 1182

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 segs, 2339354128 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

EST: *
1: gb_esc1: *
2: gb_esc2: *
3: gb_esc3: *
4: gb_hic: *
5: gb_esc4: *
6: gb_esc5: *
7: gb_esc6: *
8: gb_esc7: *
9: gb_gsa1: *
10: gb_gsa2: *
11: gb_gsa3: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	53.8	4.6	957	10	CM937254 TCBL6.1.H
C 2	45.6	3.9	700	10	BZ017393 oeb41a05-
C 3	44.6	3.8	997	10	CNS0005TE
C 4	44	3.7	1101	10	CNS0181N
C 5	43.8	3.7	807	10	CG843147
C 6	43.6	3.7	932	10	CG711559
C 7	43.2	3.7	969	1	A1557075
C 8	42.2	3.6	897	10	CZ545953
C 9	42.2	3.6	930	10	CNS00023
C 10	42	3.6	253	1	AU264198
C 11	42	3.6	253	6	CF355747
C 12	41.8	3.5	490	5	BU495620
C 13	41.8	3.5	556	5	BO452062
C 14	41.6	3.5	500	10	AU087992
C 15	41.6	3.5	610	10	CM989850
C 16	41.6	3.5	760	7	CK110580
C 17	41.6	3.5	786	8	DN495638
C 18	41.4	3.5	1035	11	CNS04JPG
C 19	41.4	3.5	1235	8	DN705852
C 20	41.2	3.5	984	10	CNS00418
C 21	41.2	3.5	987	10	CNS00418
C 22	41.2	3.5	1181	10	AG365676

23	41	3.5	378	7	CO840884	LM_GLS_00
24	41	3.5	523	10	CL375123	RC14_47
25	41	3.5	774	10	AG519237	Mus muscu
26	41	3.5	821	2	BG622779	602647567
27	41	3.5	2035	10	CL082770	CH216-171
28	40.8	3.5	437	6	CA994326	P05-39_P8
29	40.8	3.5	752	2	BE876734	601488303
30	40.8	3.5	813	7	CO085222	GR_Ea02B
31	40.6	3.5	1101	10	CNS0039G	AL063921 Drosophi
32	40.6	3.5	922	10	CL470900	SA1L_150
33	40.6	3.4	1020	6	CD050222	AGENCOURT
34	40.6	3.4	1101	10	CNS0039V	AL063936 Drosophi
35	40.6	3.4	1467	10	AG280522	AG280522 Mus muscu
36	40.4	3.4	336	1	AA014816	AA014816 mh23a06.r
37	40.4	3.4	669	10	CZ831413	OC_Ba021
38	40.2	3.4	580	5	CS85856	CS85856 Mus muscu
39	40.2	3.4	676	5	C90920	C90920 C90920 Dict
40	40.2	3.4	743	10	CE684613	CE684613 LigR-g88
41	40.2	3.4	883	9	AZ671124	AZ671124 ENTP83TF
42	40.2	3.4	885	9	AZ685382	AZ685382 ENTHA38TF
43	40.2	3.4	890	9	AZ541219	AZ541219 ENTGB47TR
44	40.2	3.4	898	9	BH139714	BH139714 ENTMY38TR
45	40.2	3.4	913	9	BH135413	BH135413 ENTMX47TF

ALIGNMENTS

RESULT 1
CM937254/c 957 bp DNA linear. GSS 20-DEC-2004
LOCUS TCBL6.1_H02_SP6 Tribolium BAC library Tribolium castaneum genomic,
DEFINITION genomic survey sequence.
ACCESSION CM937254
VERSION CM937254.1 GI:56719867
KEYWORDS GSS.
SOURCE Tribolium castaneum (red flour beetle)
ORGANISM Tribolium castaneum
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia;
Tenebrionidae; Tribolium.
1 (bases 1 to 957)
REFERENCE Savaard, J. and Tautz, D.
AUTHORS Tribolium castaneum BAC-ends sequencing project
JOURNAL Unpublished (2003)
COMMENT Contact: Savaard, J.
Abteilung fur Evolutionsgenetik, AG Tautz
Institut fur Genetik, Universitat zu Koln
Weyertal 121, 50931 Koln, Germany
Tel: 49 221 470 6911
Fax: 49 221 470 5975
Email: savaard@uni-koeln.de
Class: BAC ends.

FEATURES
source Location/Qualifiers
1..957
/organism="Tribolium castaneum"
/mol_type="genomic DNA"
/strain="GA-2"
/db_xref="taxon:7070"
/clone_lib="Tribolium BAC library"
/note="Vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI;
Library constructed by Exelixis Inc."

ORIGIN
Query Match 4.6%; Score 53.8; DB 10; Length 957;
Best Local Similarity 25.0%; Pred. No. 0.003;
Matches 142; Conservative 0; Mismatches 425; Indels 0; Gaps 0;

QY	48	AATATTATGAAAAAGCAGCATCAGAAAAATTGATGCCCTCCCTCGTGGGTGAATAT	107
DB	590	AA	531
QY	108	TAAAGCATCCCGGAGAAATCGACTGTGACAGGCTGACACCGAATATTAAGAGC	167

[illegible]

```

/organism="Brassica oleracea"
/mol_type="genomic DNA"
/db_xref="taxon:3712"
/clone_id="B.oleracea002"
/notes="Vector: pOTw13; Whole genome shotgun library from flowering buds. DNA was purified from a crude nuclear prep using Brassica oleracea TO1000DHS buds provided by Thomas Osborn at the University of Wisconsin. Genomic

```

DNA was provided by Pablo Rabinowicz (CSHL) and the shotgun library prepared at Washington University Genome Sequencing Center."

	Query Match	3.9%	Score 45.6;	DB 9;	Length 700;	
	Best Local Similarity	54.9%;	Pred. No. 0.5;			
	Matches	90;	Conservative	0;	Mismatches 74;	Indels 0; Gaps 0;
QY	483	AATTCCTTATATAGCCATCTTGTCATTACGATAGTGAATATACCGAGTTCTTTAAACA	542			
Db	244	AATTAATAATATATGAATTATATATATATATATATGAATTAACCAATTTTTTTTTAAAAA	185			
QY	543	AAAAGCAAAGAAGCAAAACAAAGTCATGTGGCATTAACCAAAAATTACAGATGA	602			
Db	184	AATGAAAAAAGTTCATTBAGCATTAAACATCTTACTTAAAAATGTAAAAATATATATATPA	125			
QY	603	GTAATTACGTAATTTTGATACATGATATCCCTGATAACGCTT	646			
Db	124	GTAATTACATATTTTAAAAAAAATGGGAAAAGTACATAAAAATT	81			

RESULT	3
CNS005TE/c	
LOCUS	
DEFINITION	CNS005TE 997 bp DNA linear GSS 03-JUN-1999
	Drosophila melanogaster genome survey sequence TEST3 end of BAC #
	BACR1XK22 of RP11-88 library from Drosophila melanogaster (fruit
	fly), genomic survey sequence.
ACCESSION	AL060767
VERSION	GI:4943573
KEYWORDS	GSS.
SOURCE	Drosophila melanogaster (fruit fly)
ORGANISM	Drosophila melanogaster

REFERENCE
1 (bases 1 to 997)

COMMENT	AUTHORS TITLE JOURNAL
<p>Genoscope. Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : segret@genoscope.cns.fr)</p> <p>- Web : www.genoscope.cns.fr/</p> <p>Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org/TheBDGP/Drosophila_melanogaster_BAC_library was prepared by Kazuhiro Oosagawa and Aaron Mamosser at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPC1-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the library p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.</p>	

```
FEATURES
  source
    1..997
      location/Qualifiers
        organism="Drosophila melanogaster"
        mol_type="genomic DNA"
        db_xref="taxon:7227"
        clone="BACCR1.2K2"
        clone_1bp="pC1-98"
        note="end : TGT3"
```

Query Match	3.88	Score 44.6	DB 10	Length 997
Best Local Similarity	18.68	Pred. No. 1		
Matches	50	Conservative 105	Mismatches 114	Indels 0
			Gaps 0	
47	GATATTATGAAAAAGCAGCATCAGAAAAATTTAGTATCCCTCCCTGCGGGGATTA	106		
07				

RESULT 5	CG843147	807 bp	DNA	linear	GSS 01-SEP-2004
LOCUS	CG843147	807 bp	DNA	linear	GSS 01-SEP-2004
DEFINITION	Yhmw3761 HW-YUEAC Bos taurus genomic clone HW-YUEAC2-188-2-C04-RP2				
ACCESSION	CG843147				
VERSION	CG843147.1	GI:51788723			
KEYWORDS	GSS.				
SOURCE	Bos taurus (cow)				
ORGANISM	Bos taurus				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.				
	1 (bases 1 to 807)				
REFERENCE	Choi, I.				
AUTHORS	Genomic sequences from Korean Cattle (Hanwoo) blood				
TITLE	Unpublished (2003)				
JOURNAL	Contact: Inho Choi				
COMMENT	Molecular Biology Yeungnam University 214-1, Dae-dong Gyeongsan, Korea 712-749 Tel: 82 53 810 2933 Fax: 82 53 816 3637 Email: inhochoi@yumail.ac.kr Seq primer: RP2 Reverse Class: BAC ends.				
FEATURES	Location/Qualifiers				
source	1..807				
	/organism="Bos taurus"				
	/mol_type="genomic DNA"				
	/strain="Korean Cattle (Hanwoo)"				
	/db_xref="taxon:9913"				
	/clone="HW-YUEAC2-188-2-C04-RP2"				
	/sex="Male"				
	/tissue_type="Blood"				
	/cell_type="Leucocyte"				
	/dev_stage="Adult"				
	/clone_id="HW-YUEAC"				
	/note="Vector: pindigobAC-5"				
ORIGIN					
	Query Match	3.7%	Score 43.8;	DB 10;	Length 807;
	Best Local Similarity	62.2%	Pred. No. 1.6;		
	Matches	69;	Conservative	0;	Mismatches 42; Indels 0; Gaps 0;
QY	160 ATAGAGCGCGGTTCAGAGACTGTTGGAAAAACAACAGAAACAGATGCAGAGATGTTAC	219			
DB	690 AATTAAGATGTTTCAACACCTGTTGGAGAGCTATGGAAACATTTTCAAGGCTAGTCTC	749			
QY	220 TGGAGACACCCGAGAGAAAAGCCGATTAAACGGAACCTCAGGGGCCGAAT	270			
DB	750 CTGGAATTAAGATTAAGAAAGGCGTGAATTCATGGAGGCACAGAGGCTAAAT	800			
RESULT 6	CG771559	932 bp	DNA	linear	GSS 29-OCT-2003
LOCUS	CG771559	932 bp	DNA	linear	GSS 29-OCT-2003
DEFINITION	TC842.1 B10 SP6 Tribolium BAC library Tribolium castaneum genomic,				
ACCESSION	CG771559				
VERSION	CG771559.1	GI:38024738			
KEYWORDS	GSS.				
SOURCE	Tribolium castaneum (red flour beetle)				
ORGANISM	Tribolium castaneum				
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia; Tenebrionidae; Tribolium.				

REFERENCE 1 (bases 1 to 932)
AUTHORS Savard, J. and Tautz, D.
TITLE Tribolium castaneum BAC-ends sequencing project
JOURNAL Unpublished (2003)
COMMENT Contact: Savard, J.
Abteilung fur Evolutionsgenetik, AG Tautz
Institut fur Genetik, Universitat zu Koln
Weyertal 121, 50931 Koln, Germany
Tel: 49 221 470 6911
Fax: 49 221 470 5975
Email: savard@uni-koeln.de
Class: BAC ends.

FEATURES
source Location/Qualifiers
1..932
/organism="Tribolium castaneum"
/mol_type="genomic DNA"
/strain="Ga-2"
/db_xref="taxon:7070"
/clone_lib="Tribolium BAC library"
/note="Vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI;
Library constructed by Exelixis Inc."

ORIGIN
Query Match 3.7%; Score 43.6; DB 10; Length 932;
Best Local Similarity 56.2%; Pred. No. 1.9;
Matches 82; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

Db 943 GAAGCAAGTGAATCGATCTTAAATCGGTGAACACAGAACCGATGTAAGCAAGTGAAT 1002
359 GCAGAACTGAATATTATGAATAATGCTCAAAAACCGAACCAAGTGAAGAGTAGA 418

Qy 1003 GAAGCAAAATTCCTCGACAGCAAACTGATGACCTAGAGTAAAGTCAACGATGAAT 1062
419 GAGGAAAAAGTTAAAGAGTCAAGAGAAAAAGTCAAGAGTGAAGAGAGAACT 478

Db 1063 ATGGAATGTTAAAGAAAAAGATGA 1088
479 AAGGACTGAACAGAAAAAGTTCA 504

Qy
Db

RESULT 7
A1557075 969 bp mRNA linear EST 09-AUG-1999
LOCUS PT2.1.13 C03.r tumor2 Homo sapiens cDNA 3', mRNA sequence.
DEFINITION A1557075_1
ACCESSION A1557075.1 GI:4489438
VERSION A1557075.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1 (bases 1 to 969)
Huang, G.M., Ng, W.L., Faras, J., He, L., Liang, H.A., Gordon, D., Yu, J.
and Hood, L.
Prostate cancer expression profiling by cDNA sequencing analysis
Genomics 59 (2), 178-186 (1999)
10409429
Contact: Guyang Matthew Huang
Leroy Hood
University of Washington
Department of Molecular Biotechnology, Box 357730, University of
Washington, Seattle, WA 98195
Tel: 5106280100
Fax: 5106280108
Email: huanggm@yahoo.com.
Location/Qualifiers
1..969
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_lib="tumor2"
/note="Organ: Prostate; Vector: pBluescript; Directional

ORIGIN
Query Match 3.7%; Score 43.2; DB 1; Length 969;
Best Local Similarity 43.0%; Pred. No. 2.4;
Matches 114; Conservative 0; Mismatches 151; Indels 0; Gaps 0;

Db 48 AATTTTAAAAACGACGACATCAGAAAAACATTGATGCCCTCCCTGTTGGGGAAATAT 107
531 AAGGAAATNA 590

Qy 108 TAACGCATCCGCGGAAATTCGACTGTGACAGGCTGACACCGAATATTAAATAGAGC 167
591 AAAAAAAAAAAGAAANTTAACCCCAAAAAAAAAAAAAAAAAAATATTAATAAAGN 650

Db 168 CGGTTACAGACTGTTGAAAAACACAGACAGATGCAAGATGATTTACTGGGACAC 227
651 GGAANNCAAAACAAAGAACAGAAAAAAGAAAAAAGNAAAAATAGAAAAA 710

Qy 228 ACCGAGAAAAACCCGTTTAAACGAAACCTCAGGGCCGAATATCGAGAGCAAAAAA 287
711 AAATTGAAAAAGAAAAAGAAAGAAAGAAAGGCAAGGAGAAATTAACAGAAAGAA 770

Db 288 AGTCATGAGAGAAACAGTAGAAGA 312
771 AAATTAAAAAANAAAAAANAANA 795

Qy
Db

RESULT 8
C2545953 897 bp DNA linear GSS 13-MAY-2005
LOCUS SRAA-rad65a06.b1 Strongyloides ratci whole genome shotgun library.
DEFINITION (SRAGSS 004) Strongyloides ratci genomic, genomic survey sequence.
ACCESSION C2545953
VERSION C2545953.1 GI:64679422
KEYWORDS GSS.
SOURCE Strongyloides ratci
ORGANISM Strongyloides ratci
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida;
Panagrolaimoidea; Strongyloidea; Strongyloides.
1 (bases 1 to 897)
Mitreva, M., McCarter, J.P., Thompson, F., Viney, M., Pape, D.,
Ritter, E., Martin, J., Wylie, T., Dante, M., Waterston, R.H.,
Clifton, S.W. and Wilson, R.
Genome Survey sequences from the rat parasitic nematode
Strongyloides ratci
Unpublished (2005)
Contact: Mitreva M
Washington University in St. Louis
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: nematode@wustl.edu
Genomic DNA was provided by Fiona Thompson
(F.Thompson@bristol.ac.uk) and Mark Viney
(Mark.Viney@bristol.ac.uk) at the University of Bristol, Bristol,
UK.
Class: Shotgun.
Location/Qualifiers
1..897
/organism="Strongyloides ratci"
/mol_type="genomic DNA"
/strain="isofemale line ED321 heterogonic"
/db_xref="taxon:34506"
/dev_stage="infective larval stage (L3)"
/lab_host="GSI0"
/clone_lib="Strongyloides ratci whole genome shotgun
library (SRAGSS 004)"
/note="Vector: pOTW13; Site 1: BstXI; Site 2: BstXI;
Strongyloides ratci genomic DNA was randomly sheared,

ORIGIN			
Query Match	3.6%	Score 42.2;	DB 10; Length 930;
Best Local Similarity	43.8%;	Pred. No. 4.6;	
Matches 74;	Conservative 21;	Mismatches 74;	Indels 0; Gaps 0

RESULT 11	CF355747	LOCUS	DEFINITION
	CF355747	263 bp	mRNA
	1ab79f08.y1	Gastric Epithelial Progenitor	Mus musculus cDNA 5', mRNA sequence.
			linear, EST 21-AUG-200

ACCESSION CE35747
VERSION CE35747.1 GI:34022211
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 263)
AUTHORS Tidwell, R., Clifton, S., Marra, M., Hillier, L., Page, D., Martin, J., Wylie, T., Theising, B., Bowers, Y., Gibbons, M., Ritzer, E., Bennet, J., Ronko, I., Tsagarishevili, R., Belaygorod, L., Grow, A., Maguire, L., Waterston, R., and Wilson, R.
TITLE Maehu Stem cell EST Project
JOURNAL Unpublished (2002)
COMMENT Contact: Jeff Gordon and Mike Lovett
Washu, Human Genetics Division
Washington University School of Medicine
1st strand of cDNA was synthesized with reverse transcriptase and oligo(dT) beads, then cDNA was amplified by PCR using modified SMART primers. The final cDNA was amplified in PAMPI vector in annealing reaction with Uracil DNA Glycosylase (UDG). Library constructed by Y. Korshunova and M. Lovett. Library materials provided by Mills JC & Gordon JT.
FEATURES
source
1..263
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/tissue_type="Gastric Epithelial Progenitor"
/dev_stage="adult"
/lab_host="DHSalpha"
/clone_lib="Gastric Epithelial Progenitor"
/note="Vector: pBluescript SK plus; Site 1: EcoRI; Site 2: XhoI; Library was constructed by Debopam Chakrabarti. Total RNA samples were isolated from mixed stage saponin (0.1%-lysed P. falciparum 3D7 infected erythrocytes by the acidic guanidinium-phenol chloroform method. The poly A+ RNA was isolated by the polyAT-Tract RNA isolation system (Promega, WI) using streptavidin Magnosphere particles. Directional cDNA libraries were constructed by oligo d(T) priming of poly(A)+ RNA (5mg) into EcoRI and XhoI sites of 1 ZapII vector using the Zap cDNA synthesis kit (Stratagene, CA). The average size of the cDNA inserts in the library was between 1.0 and 1.5kb. Clones were mass excised using the ExAassist helper phage (Stratagene), the phagemid was precipitated with PEG 8000 and extracted into phenol/chloroform. Phagemid DNA was electroporated into DH10B cells."

ACCESSION BU495620
VERSION BU495620.1 GI:22791814
KEYWORDS EST.
SOURCE Plasmodium falciparum 3D7
ORGANISM Plasmodium falciparum 3D7
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 490)
AUTHORS Tang, K., Cole, R., Chakrabarti, D., Haywood, R., Clifton, S., Page, D., Marra, M., Hillier, L., Martin, J., Wylie, T., Dante, M., Theising, B., Bowers, Y., Gibbons, M., Ritzer, E., Bennet, J., Jentes, E., Ronko, I., Tsagarishevili, R., Belaygorod, L., Franklin, C., Carr, L., Grow, A., Maguire, L., Richey, J., Madkins, J., Kennedy, S., Levinso, D., Waterston, R., Wilson, R., and Sibley, D.
TITLE Washu Plasmodium EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: L. David Sibley
Washu Plasmodium EST Project
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
Library was constructed by Debopam Chakrabarti DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: L. David Sibley (sibley@wustl.wustl.edu), Washington University
Trace considered overall poor quality
Seq primer: -400P from Gibco
High quality sequence stop: 1.
FEATURES
source
1..490
/organism="Plasmodium falciparum 3D7"
/mol_type="mRNA"
/db_xref="taxon:36329"
/lab_host="DH10B (GeneHog, Invitrogen, Inc.)"
/clone_lib="Plasmodium falciparum 3D7 asexual cDNA"
/note="Vector: pBluescript SK plus; Site 1: EcoRI; Site 2: XhoI; Library was constructed by Debopam Chakrabarti. Total RNA samples were isolated from mixed stage saponin (0.1%-lysed P. falciparum 3D7 infected erythrocytes by the acidic guanidinium-phenol chloroform method. The poly A+ RNA was isolated by the polyAT-Tract RNA isolation system (Promega, WI) using streptavidin Magnosphere particles. Directional cDNA libraries were constructed by oligo d(T) priming of poly(A)+ RNA (5mg) into EcoRI and XhoI sites of 1 ZapII vector using the Zap cDNA synthesis kit (Stratagene, CA). The average size of the cDNA inserts in the library was between 1.0 and 1.5kb. Clones were mass excised using the ExAassist helper phage (Stratagene), the phagemid was precipitated with PEG 8000 and extracted into phenol/chloroform. Phagemid DNA was electroporated into DH10B cells."

Db 223 TAAAGTATGAGCAAGAAATTAATTAAGTAAAGAGAAATTTAAAGAAATGA 282
Qy 486 TCCTTATATGAGCCATCTTGTTCATTACAGATAGTAATATACCGAGTTCTTTAAACAAA 545
Db 283 TACTGAAATTAAGTAAGTAAGTATGAGCAAGAAATTAATTAAGTAAAGTAAAGAA 342
Qy 546 AGCAAAAGACGAAACACAAAAGTCATCTTCTGTAATCAAAAATTACGAT 600
Db 343 AATTGAAAACACTAGAAAGAAATTAAGAAATTAATTTCTGAAATTTAAAGAT 397

RESULT 13
BQ452062 556 bp mRNA linear EST 29-MAY-2002
LOCUS PFESToa91e03.y1 Plasmodium falciparum 3D7 asexual cDNA Plasmodium
DEFINITION falciparum 3D7 cDNA 5' similar to TR:Q06167 Q06167 ANTIGENIC
PROTEIN PFEMP2 ; mRNA sequence.
BQ452062
BQ452062.1 GI:21255174
SOURCE Plasmodium falciparum 3D7
ORGANISM Plasmodium falciparum 3D7
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
1 (bases 1 to 556)
Tang,K., Cole,R., Chakrabarti,D., Haywood,R., Clifton,S., Pape,D.,
Marras,M., Hillier,L., Martin,J., Wylie,T., Dente,M., Theising,B.,
Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Jentes,E., Ronko,I.,
Teagarishvili,R., Belaygorod,L., Franklin,C., Carr,L., Grow,A.,
Maguire,L., Richey,J., Madkins,J., Kennedy,S., Levinso,D.,
Waterston,R., Wilson,R. and Sibley,D.
WashU Plasmodium EST Project
Unpublished (2001)
Contact: L. David Sibley
MashU Plasmodium EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
Library was constructed by Debopam Chakrabarti DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: L. David Sibley
(lsibley@bcm.tmc.edu), Washington University
Seg primer: -40UP from Gibco
High quality sequence stop: 420.
Location/Qualifiers
1..556
/organism="Plasmodium falciparum 3D7"
/mol_type="mRNA"
/db_xref="taxon:36329"
/clone_lib="Plasmodium falciparum 3D7 asexual cDNA"
/note="Vector: pBluescript SK plus; Site_1: EcoRI; Site_2:
XhoI; Library was constructed by Debopam Chakrabarti.
Total RNA samples were isolated from mixed stage
saponin(0.1%) lysed P. falciparum 3D7 infected
erythrocytes by the acidic guanidium-phenol chloroform
method. The poly A+ RNA was isolated by the polyAT-Tract
mRNA isolation system (Promega, WI) using streptavidin
MagneSphere particles. Directional cDNA libraries were
constructed by oligo d(T) priming of poly(A)+ RNA (5mg)
into EcoRI and XhoI sites of 1 ZapII vector using the Zap
cDNA synthesis kit (Stratagene, CA). The average size of the
cDNA inserts in the library was between 1.0 and 1.5kb.
Clones were mass excised using the EXASist helper phage
(Stratagene), the phagmids were preclipped with Pst
8000 and extracted with phenol/chloroform. Phagemid DNA
was electroporated into DH10B cells."

ORIGIN
Query Match 3.5%; Score 41.8; DB 5; Length 556;
Best Local Similarity 47.0%; Pred. No. 5.2;

Matches 167; Conservative 0; Mismatches 182; Indels 6; Gaps 1;
Qy 246 TAAACGAAACCTCAGAGGCGCCGAATATCGAGACAAAGAAAGTCATGAGAAACAGT 305
Db 56 TAAAGTAAAGGCGCCAGAAATTAATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 115
Qy 306 AGAAAGAGGCGCTAACCGCATTTGTCAGTAAATGCAATGCGCTGTTAACCAATTTCAAT 365
Db 116 AGAAGAAAGAAATTAAGAAATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 175
Qy 366 CATCTTTGAGAGAACTTTCTGACAGCCCAATATCGCGCTCATTTGATGTTTGAAGA 425
Db 176 AATACTGAAAGAGTAA-----AAGAAGAAATTAAGAAATTAAGTAAAGTAAAGTAA 229
Qy 426 CCATATGATGTCATGAGGCGCCAGCGCTGATGAATTTGAGAGCGTTACCGCTACAT 485
Db 230 TAAAGTATGAGCAAGAAATTAATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 289
Qy 486 TCCTTATATGAGCCATCTTGTTCATTACAGATAGTAATATACCGAGTTCTTTAAACAAA 545
Db 290 TACTGAAATTAAGTAAGTAAGTATGAGCAAGAAATTAATTAAGTAAAGTAAAGTAA 349
Qy 546 AGCAAAAGACGAAACACAAAAGTCATCTTCTGTAATCAAAAATTACGAT 600
Db 350 AATTGAAAACACTAGAAAGAAATTAAGAAATTAATTTCTGAAATTTAAAGAT 404

RESULT 14
LOCUS AU087992/c 500 bp mRNA linear EST 08-MAR-2005
DEFINITION AU087992 Sugano Malaria cDNA library Plasmodium falciparum 3D7 cDNA
clone XPFn6290, mRNA sequence.
AU087992
AU087992.1 GI:12390133
SOURCE Plasmodium falciparum 3D7
ORGANISM Plasmodium falciparum 3D7
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
1 (bases 1 to 500)
Watanabe,J., Sasaki,M., Suzuki,Y. and Sugano,S.
FULT-malaria: a database for a full-length enriched cDNA library
from human malaria parasite, Plasmodium falciparum
Nucleic Acids Res. 29 (1), 70-71 (2001)
1125052
JOURNAL Contact: Junichi Watanabe
PUBMED Institute of Medical Science
COMMENT The University of Tokyo, Department of Parasitology
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Tel: 81-3-5449-5283
Fax: 81-3-5449-5416
Email: jwatanab@ims.u-tokyo.ac.jp,
URL: http://fuliml.ims.u-tokyo.ac.jp
Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,A. and
Sugano,S. Construction and characterization of a full
length-enriched and a 5'-end-enriched cDNA library Gene 200 (1-2),
149-156 (1997).
Location/Qualifiers
1..500
/organism="Plasmodium falciparum 3D7"
/mol_type="mRNA"
/isolate="3D7"
/db_xref="taxon:36329"
/clone_lib="XPFn6290"
/dev_stage="erythrocytic stage"
/clone_lib="Sugano Malaria cDNA library"

ORIGIN
Query Match 3.5%; Score 41.6; DB 1; Length 500;
Best Local Similarity 54.6%; Pred. No. 5.7; Indels 6; Gaps 0;
Matches 83; Conservative 0; Mismatches 69;

Qy 458 AATTTGCAAGAGCGTTACCGCTACAAATCTTATATAGCGCATCTTGTTCATTACAGATA 517

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:34:16 ; Search time 746.204 Seconds
(without alignments)
10556.990 Million cell updates/sec

Title: US-10-789-164-1

Perfect score: 1182

Sequence: 1 atgggcctggtactcattat.....agtaagcagctcgtaagc 1182

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq_21:*

- 1: Geneseq1980s:*
- 2: Geneseq1990s:*
- 3: Geneseq2000s:*
- 4: Geneseq2001as:*
- 5: Geneseq2001bs:*
- 6: Geneseq2002as:*
- 7: Geneseq2002bs:*
- 8: Geneseq2003as:*
- 9: Geneseq2003bs:*
- 10: Geneseq2003cs:*
- 11: Geneseq2003ds:*
- 12: Geneseq2004as:*
- 13: Geneseq2004bs:*
- 14: Geneseq2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1182	100.0	1182	10	ADA44791 Bacillus
2	1182	100.0	1182	12	AD0007641 B. subtili
3	1182	100.0	1182	13	ADV65732 B. subtili
4	1182	100.0	6536	10	ADA44796
5	1171.2	99.1	1182	4	AAP82255
6	1171.2	99.1	3045	4	AAP82254
7	548.4	46.4	819	6	ABK75403
8	338.6	28.6	1164	13	ABN91998
9	338.6	28.6	1164	13	AD01519
10	338.6	28.6	3340	4	AAS4379
11	313.2	26.5	1212	9	ADB86113
12	313.2	26.5	110000	9	ADB12064
13	294.2	24.9	560	13	ADR48536
14	246.2	20.8	861	4	AAS4379
15	181	15.3	309	6	ABK75876
16	52.6	4.5	2000	11	ADA71938
17	44.6	3.8	2000	11	ACU37108
18	44.4	3.8	1338	8	ACA42892
19	42.6	3.6	600	13	ACN56103

ALIGNMENTS

C	20	42.2	3.6	434	5	ADL44027	ADL44027 Human ova
C	21	42	3.6	12507	6	ABL32299	ABL32299 Human imm
C	22	41	3.5	6000	6	ABL34305	ABL34305 Human imm
C	23	40.8	3.5	2000	8	ADA71938	ADA71938 Rice gene
C	24	40.2	3.4	112	13	ADR48527	ADR48527 capB tag
C	25	40	3.4	7312	6	ABL33815	ABL33815 Human imm
C	26	40	3.4	7312	6	ABL70402	ABL70402 Chemical1
C	27	40	3.4	7312	6	AAS61350	AAS61350 Human gen
C	28	39.8	3.4	11944	6	ABL34186	ABL34186 Human imm
C	29	39.6	3.4	1047	6	ABN93213	ABN93213 Staphyloc
C	30	39.6	3.4	1047	13	AD04156	AD04156 Staphyloc
C	31	39.6	3.4	1546	10	AD087518	AD087518 Human GPC
C	32	39.6	3.4	4642	4	AAS4726	AAS4726 S. epider
C	33	39.2	3.3	13133	6	ABK31230	ABK31230 Signal tr
C	34	39.2	3.3	3095	2	AAC03875	AAC03875 Sequence
C	35	38.8	3.3	7057	7	AAS46601	AAS46601 Tumour su
C	36	38.8	3.3	7057	6	ABL33848	ABL33848 Human imm
C	37	38.8	3.3	7057	6	ABL34594	ABL34594 Human met
C	38	38.8	3.3	7057	6	ABL70407	ABL70407 Chemical1
C	39	38.8	3.3	7057	6	AAS61354	AAS61354 Human gen
C	40	38.8	3.3	7057	7	ADS99855	ADS99855 Bistulphit
C	41	38.6	3.3	1008	8	ACA46836	ACA46836 Prokaryot
C	42	38.6	3.3	1011	4	AAS53908	AAS53908 S. epider
C	43	38.6	3.3	5311	6	ABL33019	ABL33019 Human imm
C	44	38.6	3.3	103929	3	AAP22287	AAP22287 BAC conta
C	45	38.4	3.2	511	13	ACN45990	ACN45990 Cotton pr

RESULT 1	ADA44791	standard; DNA; 1182 BP.
ID	ADA44791	
XX	AC	ADA44791;
XX	AC	20-NOV-2003 (first entry)
DE	XX	Bacillus subtilis poly-gamma-glutamate synthetase complex psbB gene.
DE	XX	Cell surface; expression vector; microbial;
KW	XX	poly-gamma-glutamate synthetase; pgsbA complex; surface expression;
KW	XX	Gram-positive bacterium; Gram-negative bacterium; enzyme; antigen;
KW	XX	antibody; attachment protein; adsorption protein; vaccine; pgsB; gene;
OS	XX	Bacillus subtilis; variety chungkookjang.
XX	XX	WO2003014360-A1.
PN	XX	20-FEB-2003.
PD	XX	09-AUG-2002; 2002WO-KR001522.
PF	XX	10-AUG-2001; 2001KR-00048373.
PR	XX	(BIOL-) BIOLEADERS CORP.
PA	XX	(MDMP-) MD LAB CO LTD.
PA	XX	Sung M, Hong S, Lee J, Jung C, Kim C, Soda K, Ashiuchi M;
PI	XX	WPI; 2003-256589/25.
DR	XX	New expression vector containing gene(s) that encode a poly-gamma-
PT	XX	glutamate synthetase complex, useful for producing proteins (e.g.
PT	XX	vaccines or enzymes) on the microbial surface of Gram-positive and/or
PT	XX	Gram-negative bacteria.
PS	XX	Claim 3; Page 109-110; 122pp; English.
CC	XX	The invention relates to a vector for expression of a target protein on a
CC	XX	microbial cell surface. The vector of the invention comprises either one

or more than two genes of the *Bacillus subtilis* poly- γ -glutamate synthetase complex (pgsBCA) to facilitate microbial surface expression of the target protein. The pgsBCA gene complex comprises the pgsB, pgsB and pgsA genes and is normally expressed in the outer membrane of *Bacillus subtilis*. The vector can be transformed into either Gram-positive or Gram-negative bacteria (e.g., *Escherichia coli*), and can be used for the surface expression of various proteins of interest such as enzymes, antigens, antibodies, attachment proteins or adsorption proteins. CC proteins recombinantly produced using the vector of the invention can be used as, for example, vaccines or enzymes. The present sequence CC represents the *Bacillus subtilis* pgsBCA complex gene pgsB, which is specifically claimed for use in the vector of the invention.

Sequence 1182 BP; 396 A; 237 C; 274 G; 275 T; 0 U; 0 Other;

Query Match 100.0%; Score 1182; DB 10; Length 1182;
Beet Local Similarity 100.0%; Pred. No. 0;
Matches 1182; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Oy 1 ATGGGCTGTACTCATTTATAGCTGTGTCATCTGTCATCGAATATTGAAAA 60
Db 1 ATGGGCTGTACTCATTTATAGCTGTGTCATCTGTCATCGAATATTGAAAA 60
Oy 61 CGACGACATCAGAAAAACATTGATGCCCTCCCTGTTCCGGTGAATATTACGGCATCCG 120
Db 61 CGACGACATCAGAAAAACATTGATGCCCTCCCTGTTCCGGTGAATATTACGGCATCCG 120
Oy 121 GGAATATGACTGTGACAAAGCTGACACCGGAATTTATAGAACCCGGTTACAAGCT 180
Db 121 GGAATATGACTGTGACAAAGCTGACACCGGAATTTATAGAACCCGGTTACAAGCT 180
Oy 181 GTTGGAAAAACAACAGAAAGATGCAAGATGATTTACTGGGACACACCGAGAGAAAG 240
Db 181 GTTGGAAAAACAACAGAAAGATGCAAGATGATTTACTGGGACACACCGAGAGAAAG 240
Oy 241 CCGATTTAAACGAAACCTCAGGGGCCGAATTCGAGAGACAAAAAGAGTCATGAGAGA 300
Db 241 CCGATTTAAACGAAACCTCAGGGGCCGAATTCGAGAGACAAAAAGAGTCATGAGAGA 300
Oy 301 ACAGTAGAAAAAGGGGCTTACGCCATTTGTCACTGATGATGAGCTTTAACCCAGATTAT 360
Db 301 ACAGTAGAAAAAGGGGCTTACGCCATTTGTCACTGATGATGAGCTTTAACCCAGATTAT 360
Oy 361 CAATATCATCTTTCAGGAGAACTTCTGACAGGCAATATCGGGTCATTTGTAATGTTTA 420
Db 361 CAATATCATCTTTCAGGAGAACTTCTGACAGGCAATATCGGGTCATTTGTAATGTTTA 420
Oy 421 GAAGACATATGATGTATGGGGCCGACGCTTGAATGAATTTGCAAGACGTTTACCGCT 480
Db 421 GAAGACATATGATGTATGGGGCCGACGCTTGAATGAATTTGCAAGACGTTTACCGCT 480
Oy 481 ACAATCTCTTATATAGGCGATCTTGTCTATACAGATAGTGAATATCCGAGTTCTTTTAA 540
Db 481 ACAATCTCTTATATAGGCGATCTTGTCTATACAGATAGTGAATATCCGAGTTCTTTTAA 540
Oy 541 CAAAAAGCAAAAGCAAGAAACAAAGATCATCTGTCATGATCACTAAAAATTACAGAT 600
Db 541 CAAAAAGCAAAAGCAAGAAACAAAGATCATCTGTCATGATCACTAAAAATTACAGAT 600
Oy 601 GAGTATTTACGTAATTTTGAATACATGATATTCCTGTAATACGCTTCTGCGCTGGGT 660
Db 601 GAGTATTTACGTAATTTTGAATACATGATATTCCTGTAATACGCTTCTGCGCTGGGT 660
Oy 661 GTGGCTCAAGCACTGCGCATTTGACAGAAAGACATTTAAGGAATGCTGAATCGCGG 720
Db 661 GTGGCTCAAGCACTGCGCATTTGACAGAAAGACATTTAAGGAATGCTGAATCGCGG 720
Oy 721 CCAGATCCGGGAGCAATGAAATCTTCCGCTGATCAGTCGAGGAGAGCTGGGACATTT 780
Db 721 CCAGATCCGGGAGCAATGAAATCTTCCGCTGATCAGTCGAGGAGAGCTGGGACATTT 780
Oy 781 GTTAATGGGTTTCCGCAACGACGCTTCTTACTTTGAATATATGAAACGTGTAAAA 840
Db 781 GTTAATGGGTTTCCGCAACGACGCTTCTTACTTTGAATATATGAAACGTGTAAAA 840

```

```

Db 781 GTTAATGGGTTTCCGCAACGACGCTTCTTACTTTGAATATATGAAACGTGTAAAA 840
Oy 841 GAAATCGTTATCCGACCGATGATTCGATCATCATCAATGACCGCGAGACCGTGTG 900
Db 841 GAAATCGTTATCCGACCGATGATTCGATCATCATCAATGACCGCGAGACCGTGTG 900
Oy 901 GATCGACACAGCAATTCGCAATGACGATATTCGCTTATATTTGAAGCAATGATC 960
Db 901 GATCGACACAGCAATTCGCAATGACGATATTCGCTTATATTTGAAGCAATGATC 960
Oy 961 TTAATCGGTGAACACAGAACCGATGATTAAGCCCTATGAAAGGCAAAATTCCTGCA 1020
Db 961 TTAATCGGTGAACACAGAACCGATGATTAAGCCCTATGAAAGGCAAAATTCCTGCA 1020
Oy 1021 GACAACGTGATGACCTAGATATATGATCAACAGATGAATTTGAAATTTGTTAAAGAA 1080
Db 1021 GACAACGTGATGACCTAGATATATGATCAACAGATGAATTTGAAATTTGTTAAAGAA 1080
Oy 1081 AGAATGCAACAACCGTGCATATATATGCGTGGCAATATTCATGTGCGGACAGACCTTTA 1140
Db 1081 AGAATGCAACAACCGTGCATATATATGCGTGGCAATATTCATGTGCGGACAGACCTTTA 1140
Oy 1141 ATTGAAAAATCCACGAATACAGGTAAAGCACTCGTAAGC 1182
Db 1141 ATTGAAAAATCCACGAATACAGGTAAAGCACTCGTAAGC 1182

RESULT 2
AD007641
ID AD007641 standard; DNA; 1182 BP.
AC AD007641;
XX
XX 15-JUL-2004 (first entry)
DT
DE B subtilis poly-X-glutamate synthetase complex coding sequence pgsB.
DE ds; gene; enzyme; vaccine; cytostratic; pgsB; poly-X-glutamate synthetase;
KW human papilloma virus.
OS
XX
XX
XX WO2004035795-A1.
XX
XX 29-APR-2004.
XX
XX 17-OCT-2003; 2003WO-KR002163.
XX
XX 17-OCT-2002; 2002KR-00063378.
XX
XX
XX (BIOL-) BIOLEADERS CORP.
XX (KORE-) KOREA RES INST BIOSCIENCE & BIOTECHNOLOG.
XX
XX Sung M, Poo H, Lee J, Jung C, Hong S, Kim C, Park S, Pyo H;
XX WPI; 2004-348463/32.
XX
XX New vector containing pgs A-C genes encoding poly- $\gamma$ -glutamate
XX synthetase complex and an antigen protein gene of human papilloma virus,
XX useful in preparing vaccine for treating or preventing mucosal tumor,
XX e.g. cervical cancer.
XX
XX Disclosure; Page 60-61; 69pp; English.
XX
XX The present invention relates to a vector for preparing a vaccine which
XX contains one or more than two genes, i.e. pgs A-C encoding poly-X-
XX glutamate synthetase complex and an antigen protein gene of human
XX papilloma virus. The vector and microbes transformed with it are useful
XX in preparing vaccines for treating or preventing mucosal tumour, e.g.
XX cervical cancer. The present sequence is a Bacillus subtilis poly-X-
XX glutamate synthetase complex coding sequence.
XX
XX Sequence 1182 BP; 396 A; 237 C; 274 G; 275 T; 0 U; 0 Other;

```

Query Match 100.0%; Score 1182; DB 12; Length 1182;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1182; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ATGGGCTGTTACTATATAGCCCTGTGTCATCATCTGTCATGGAATATTAGAAAA 60
DB 1 ATGGGCTGTTACTATATAGCCCTGTGTCATCATCTGTCATGGAATATTAGAAAA 60
QY 61 CGACGACATCAGAAAAACATTGATGCCCTCCCTGTTGGGTTAAATTTAACGGCATCCGC 120
DB 61 CGACGACATCAGAAAAACATTGATGCCCTCCCTGTTGGGTTAAATTTAACGGCATCCGC 120
QY 121 GGAATATGACTGTGACAAAGCTGACAAACCGGAATTTATTTAGAACCGGTTACAAAGCT 180
DB 121 GGAATATGACTGTGACAAAGCTGACAAACCGGAATTTATTTAGAACCGGTTACAAAGCT 180
QY 181 GTTGGAAAAACAAACAGAAACAGATGCAAGATTTTCTGGACACACCGGAGGAAAAAG 240
DB 181 GTTGGAAAAACAAACAGAAACAGATGCAAGATTTTCTGGACACACCGGAGGAAAAAG 240
QY 241 CCGATTTAAACGGAACCTCAGGGGCGGAATATCGAGAGCAAAAAAGATCATGAGAGA 300
DB 241 CCGATTTAAACGGAACCTCAGGGGCGGAATATCGAGAGCAAAAAAGATCATGAGAGA 300
QY 301 ACAGTAGAAAGAGGGGCTTAAACGCAATGTCAGTGAATGATGCTGTTAACCCAGATTAT 360
DB 301 ACAGTAGAAAGAGGGGCTTAAACGCAATGTCAGTGAATGATGCTGTTAACCCAGATTAT 360
QY 361 CAAATCATCTTTGAGGAAGAACTTCTGCGAGGCAATATGGGGTCAATTTGTAATGTTTA 420
DB 361 CAAATCATCTTTGAGGAAGAACTTCTGCGAGGCAATATGGGGTCAATTTGTAATGTTTA 420
QY 421 GAAAGACATATGATGTCTGAGGGCCGACGCTTGAATGAATTTGCAGAACGCTTTACCGCT 480
DB 421 GAAAGACATATGATGTCTGAGGGCCGACGCTTGAATGAATTTGCAGAACGCTTTACCGCT 480
QY 481 ACAATTCCTTATTAATGGCCATCTTGTCAATTACAGATGATGATTAACGAGTTCTTTAA 540
DB 481 ACAATTCCTTATTAATGGCCATCTTGTCAATTACAGATGATGATTAACGAGTTCTTTAA 540
QY 541 CAAAAAGCAAAAGAAACGAAACAAAGATCATCTGTCATGATCACTCAAAAATTTACAGAT 600
DB 541 CAAAAAGCAAAAGAAACGAAACAAAGATCATCTGTCATGATCACTCAAAAATTTACAGAT 600
QY 601 GAGTATTTCGTAAATTTGAAATATCATGATATCCCTGATTAACGCTTCTGCGGCTGGGT 660
DB 601 GAGTATTTCGTAAATTTGAAATATCATGATATCCCTGATTAACGCTTCTCTGCGGCTGGGT 660
QY 661 GTGGCTCAAGCACTCGGCATTTGACGAAGAAACGCAATTTAAGGAATGCTGAATGCGCCG 720
DB 661 GTGGCTCAAGCACTCGGCATTTGACGAAGAAACGCAATTTAAGGAATGCTGAATGCGCCG 720
QY 721 CCGATCCGAGGACAAATGAGAAATTTCTCCGCTGATCATGTCGAGCGCTGGGCACTTT 780
DB 721 CCGATCCGAGGACAAATGAGAAATTTCTCCGCTGATCATGTCGAGCGCTGGGCACTTT 780
QY 781 GTTAAATGGGTTGGCCCAAAACGACGCTTCTTACTTTGAATATATGAAACGCTGTAATA 840
DB 781 GTTAAATGGGTTGGCCCAAAACGACGCTTCTTACTTTGAATATATGAAACGCTGTAATA 840
QY 841 GAAATCGGTTACCGGACCGATGATCCGATCATGATGAACTGCGCGACGACCGGTGC 900
DB 841 GAAATCGGTTACCGGACCGATGATCCGATCATGATGAACTGCGCGACGACCGGTGC 900
QY 901 GATCGGACACAGCAATTTGCAATGACGATTTGCTTTAATTTGAAGCAATGAACTGATC 960
DB 901 GATCGGACACAGCAATTTGCAATGACGATTTGCTTTAATTTGAAGCAATGAACTGATC 960
QY 961 TTAATCGGTGAACAAACAGAACGATGCTTAAAGCCTATGAAAGAGCAAAATTTCTGCA 1020
DB 961 TTAATCGGTGAACAAACAGAACGATGCTTAAAGCCTATGAAAGAGCAAAATTTCTGCA 1020

```

```

QY 1021 GACAAACTGCATGACCTAGATATAGTCAACAGATGAAATTTAGATTGTAAGAAA 1080
DB 1021 GACAAACTGCATGACCTAGATATAGTCAACAGATGAAATTTAGATTGTAAGAAA 1080
QY 1081 AGAATGCAACACCGTGTCAATATGAGGCTGGGCAATATTCATGTGCGCGACACCTTTA 1140
DB 1081 AGAATGCAACACCGTGTCAATATGAGGCTGGGCAATATTCATGTGCGCGACACCTTTA 1140
QY 1141 ATTGAAAAAATCCAGCAATACAGGTTAAAGCAGCTCGTAAC 1182
DB 1141 ATTGAAAAAATCCAGCAATACAGGTTAAAGCAGCTCGTAAC 1182

```

RESULT 3

ADV65732 standard; DNA; 1182 BP.

ADV65732; 10-FEB-2005 (first entry)

B. subtilis poly-gamma-glutamic acid synthase B gene SEQ ID NO:1.
 ds; poly-gamma-glutamic acid synthase B; pgsB; gene expression;
 antibiotic; antimicrobial; fungicide; cytostatic.

Bacillus subtilis.

KR2004034780-A.

29-APR-2004.

17-OCT-2002; 2002KR-00063379.

17-OCT-2002; 2002KR-00063379.

(BIOL-) BIOLEADERS CORP.

(UYCH-) UNIV CHUNSUN CO LTD.

(KORE-) KOREA RES INST BIOSCIENCE & BIOTECHNOLOG.

Boo HR, Ham GS, Hong SP, Jung CM, Kim CJ, Lee DG, Lee JS;

Park YG, Sung MH;

WPI; 2004-577380/56.

Method for surface expression of peptides p5 and ana13 using pgs bca

gene.

Disclosure; SEQ ID NO 1; 25pp; Korean.

The invention relates to a novel method for surface expression of
 peptides p5 and Ana13 using a poly-gamma-glutamic acid synthase (pgs) BCA
 gene, thereby removing a purification process of peptides p5 and Ana13,
 and using lactic acid bacteria for the surface expression, so that
 peptide antibiotics can be cheaply and stably mass-produced. An
 expression vector pKCB11B:pgsA-P5 comprises one or more genes encoding
 poly-gamma-glutamic acid synthase selected from pgsB, pgsC and pgsA, and
 a gene for dipolar peptide antibiotics having antimicrobial, antifungal
 and anticancer activities, wherein the dipolar peptide antibiotic has
 homology to the peptide p5 encoded by the nucleotide sequence set forth
 in ADV65735, or to the peptide Ana13 encoded by the nucleotide sequence
 set forth in ADV65737. The present sequence represents the B. subtilis
 poly-gamma-glutamic acid synthase B gene used in the invention.

Sequence 1182 BP; 396 A; 237 C; 274 G; 275 T; 0 U; 0 Other;

Query Match 100.0%; Score 1182; DB 13; Length 1182;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1182; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ATGGGCTGTTACTATATAGCCCTGTGTCATCATCTGTCATGGAATATTAGAAAA 60
DB 1 ATGGGCTGTTACTATATAGCCCTGTGTCATCATCTGTCATGGAATATTAGAAAA 60

```

```
QY 61 CGACGACATCAGAAAAAATTGATGCTCCCTGTTCCGGTGAAATATTAAAGGATCCGC 120
Db 61 CGAGACATCAGAAAAAATTGATGCTCCCTGTTCCGGTGAAATATTAAAGGATCCGC 120
QY 121 GGAATAATGACTGTGACAAAGCTGACAAACCGAATATTAAATGAAGCCGTTTCAAGACT 180
Db 121 GGAATAATGACTGTGACAAAGCTGACAAACCGAATATTAAATGAAGCCGTTTCAAGACT 180
QY 181 GTTGAAAAAACAACAGAAAGATGACAAAGATTTTCTGAGGACACACCGGAGGAAAAAG 240
Db 181 GTTGAAAAAACAACAGAAAGATGACAAAGATTTTCTGAGGACACACCGGAGGAAAAAG 240
QY 241 CCGATTAAACGGAAACCTCAGGGGCGGAATATCGGAGAGCAAAAAAGAGTCATGAGAGAA 300
Db 241 CCGATTAAACGGAAACCTCAGGGGCGGAATATCGGAGAGCAAAAAAGAGTCATGAGAGAA 300
QY 301 ACAGTAGAAGAGGGGCTTAACGCGATTGTCAGTGAATGATGCTGTTAAACCCAGATTAT 360
Db 301 ACAGTAGAAGAGGGGCTTAACGCGATTGTCAGTGAATGATGCTGTTAAACCCAGATTAT 360
QY 361 CAATTCATCTTTCCAGAAACCTTGTGAGGCGCAATATCGCGTCATTTGTAATGTTTAA 420
Db 361 CAATTCATCTTTCCAGAAACCTTGTGAGGCGCAATATCGCGTCATTTGTAATGTTTAA 420
QY 421 GAAGACCATATGATGATGATGGGGCCGACGCTTGATGAATTTGCAAGCGTTTAAACCGCT 480
Db 421 GAAGACCATATGATGATGATGGGGCCGACGCTTGATGAATTTGCAAGCGTTTAAACCGCT 480
QY 481 ACAATTCCTTAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540
Db 481 ACAATTCCTTAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540
QY 541 CAAAAGCAAAAGAAAGCAAAACCAAAAGTCATGCTGATGATGATGATGATGATGATGATGAT 600
Db 541 CAAAAGCAAAAGAAAGCAAAACCAAAAGTCATGCTGATGATGATGATGATGATGATGATGAT 600
QY 601 GAGTATTTACGTAATTTTGAATATACATGATGATGATGATGATGATGATGATGATGATGAT 660
Db 601 GAGTATTTACGTAATTTTGAATATACATGATGATGATGATGATGATGATGATGATGATGAT 660
QY 661 GTGGCTCAGACATCGGCAATGACGAAAGAAACAGCATTTAAAGGCAATCTGAATGCGCG 720
Db 661 GTGGCTCAGACATCGGCAATGACGAAAGAAACAGCATTTAAAGGCAATCTGAATGCGCG 720
QY 721 CCAATTCGGGAGCAATGAAATCTTCCGCTGATGATGATGATGATGATGATGATGATGATGAT 780
Db 721 CCAATTCGGGAGCAATGAAATCTTCCGCTGATGATGATGATGATGATGATGATGATGATGAT 780
QY 781 GTTAATGAGTTTGGCGCAAAAGCAAGCTTCTTCTACTTGAATATATGAAACGTGTAAAA 840
Db 781 GTTAATGAGTTTGGCGCAAAAGCAAGCTTCTTCTACTTGAATATATGAAACGTGTAAAA 840
QY 841 GAAATTCGGTTTGGCGCAAAAGCAAGCTTCTTCTACTTGAATATATGAAACGTGTAAAA 900
Db 841 GAAATTCGGTTTGGCGCAAAAGCAAGCTTCTTCTACTTGAATATATGAAACGTGTAAAA 900
QY 901 GATGAGACACAGCAATTTGCAAAATGACGTAATGCTTATATGAAAGCAAGTGAATC 960
Db 901 GATGAGACACAGCAATTTGCAAAATGACGTAATGCTTATATGAAAGCAAGTGAATC 960
QY 961 TTAATTCGGTGAACAAACAGAACCGATCGTAAAGAGCTTAAGAAAGCAAAATTCCTGCA 1020
Db 961 TTAATTCGGTGAACAAACAGAACCGATCGTAAAGAGCTTAAGAAAGCAAAATTCCTGCA 1020
QY 1021 GACAAATCGATGACCTTAAGATTAAGTCAACAGATGAATTAATGAAATGTTGTTAAAGAA 1080
Db 1021 GACAAATCGATGACCTTAAGATTAAGTCAACAGATGAATTAATGAAATGTTGTTAAAGAA 1080
QY 1081 AGAATGCAACACCGTGTATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
Db 1081 AGAATGCAACACCGTGTATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
```

```
QY 1141 ATTGAAAAAATCCACGAATACAAAGTAAAGAGCTGCTAAGC 1182
Db 1141 ATTGAAAAAATCCACGAATACAAAGTAAAGAGCTGCTAAGC 1182

RESULT 4
ADA44796
ID ADA44796 standard; DNA; 6536 BP.
XX
AC ADA44796;
XX
DT 20-NOV-2003 (first entry)
XX
DE Microbial cell surface expression vector pGNBCA, SEQ ID NO:6.
XX
KW Cell surface; expression vector; microbial;
KW poly-gamma-glutamate synthetase; Bacillus subtilis pGNBCA complex;
KW surface expression; Gram-positive bacterium; Gram-negative bacterium;
KW enzyme; antigen; antibody; attachment protein; adsorption protein;
KW vaccine; pGNBCA; cyclic; circular; ds.
XX
OS Synthetic.
OS Bacillus subtilis; variety chungkookjang.
XX
PN WO2003014360-A1.
XX
PD 20-FEB-2003.
XX
PE 09-AUG-2002; 2002MO-KR001522.
XX
PR 10-AUG-2001; 2001KR-00048373.
XX
PA (BIOL-) BIOLEADERS CORP.
PA (MDMD-) MD LAB CO LTD.
XX
PI Sung M, Hong S, Lee J, Jung C, Kim C, Soda K, Ashuchi M;
XX
DR WPI; 2003-256589/25.
XX
PT New expression vector containing gene(s) that encode a poly-gamma-
PT glutamate synthetase complex, useful for producing proteins (e.g.
PT vaccines or enzymes) on the microbial surface of Gram-positive and/or
PT Gram-negative bacteria.
XX
PS Example 1; Page 113-120; 122pp; English.
XX
CC The invention relates to a vector for expression of a target protein on a
CC microbial cell surface. The vector of the invention comprises either one
CC or more than two genes of the Bacillus subtilis poly-gamma-glutamate
CC synthetase complex (pGNBCA) to facilitate microbial surface expression of
CC the target protein. The pGNBCA gene complex comprises the pgsB, pgsB and
CC pgsA genes and is normally expressed in the outer membrane of Bacillus
CC subtilis. The vector can be transformed into either Gram-positive or Gram
CC -negative bacteria (e.g., Escherichia coli), and can be used for the
CC surface expression of various proteins of interest such as enzymes,
CC antigens, antibodies, attachment proteins or adsorption proteins.
CC Proteins recombinantly produced using the vector of the invention can be
CC used as, for example, vaccines or enzymes. The present sequence
CC represents a microbial cell surface expression vector of the invention,
CC pGNBCA.
XX
SQ Sequence 6536 BP; 1855 A; 1464 C; 1567 G; 1649 T; 0 U; 1 Other;

Query Match 100.0%; Score 1182; DB 10; Length 6536;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1182; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGGCTGTTACTCATTAATAGCTGCTGCTCATCTGATCGATCGAATATTGAAAAA 60
Db 224 ATGGGCTGTTACTCATTAATAGCTGCTGCTCATCTGATCGATCGAATATTGAAAAA 283
QY 61 CGACGACATCAGAAAAAATTGATGCTCCCTGTTCCGGTGAAATATTAAAGGATCCGC 120
```

```

Db 284 CGACGACATCAGAAAAACATTGATGCGCTCCCTGTCGGTGAATATTAAAGCGCATCCGC 343
Qy 121 GGAATAATCGACTGTGACAAAGGCTGACAAACCGGAATATTATAGAACCGGTTACAAAGCT 180
Db 344 GGAATAATCGACTGTGACAAAGGCTGACAAACCGGAATATTATAGAACCGGTTACAAAGCT 403
Qy 181 GTTGAAAAACAACGAAACAGATGCGAAGATTTACTGGAGACACACCGGAGGAAAG 240
Db 404 GTTGAAAAACAACGAAACAGATGCGAAGATTTACTGGAGACACACCGGAGGAAAG 463
Qy 241 CCGATTTAAACGGAACCTCAGGGGCGGAATATCGAGAGCAAAAAGAGTCATGAGAGA 300
Db 464 CCGATTTAAACGGAACCTCAGGGGCGGAATATCGAGAGCAAAAAGAGTCATGAGAGA 523
Qy 301 ACAGTAGAAAGAGGGGCTTAACCGCATTTGTCAGTGAATGCGCTGTTAACCGAATTAT 360
Db 524 ACAGTAGAAAGAGGGGCTTAACCGCATTTGTCAGTGAATGCGCTGTTAACCGAATTAT 583
Qy 361 CAATATCATTTTCAGAGAAACTTTCGACGGCCAAATATCGGGGTCATTTGTAATGTTTA 420
Db 584 CAATATCATTTTCAGAGAAACTTTCGACGGCCAAATATCGGGGTCATTTGTAATGTTTA 643
Qy 421 GAAGACCATATGAGATGTCATGGGGCCGAGCTTGATGAATTCAGAAACGTTTACCGCT 480
Db 644 GAAGACCATATGAGATGTCATGGGGCCGAGCTTGATGAATTCAGAAACGTTTACCGCT 703
Qy 481 ACAATTCCTTATATATGCGCATCTTGTCAATTAAGATAGTGAATATACGAGTTCTTTAA 540
Db 704 ACAATTCCTTATATATGCGCATCTTGTCAATTAAGATAGTGAATATACGAGTTCTTTAA 763
Qy 541 CAAAAAGCAAAAGAACGAAACACAAAAGTCATATGCTGATACCTCAAAAATTTAAGAT 600
Db 764 CAAAAAGCAAAAGAACGAAACACAAAAGTCATATGCTGATACCTCAAAAATTTAAGAT 823
Qy 601 GAGTATTTAGTATTTTGAATATCATGATGATTCCTGATTAACGCTTCTGCGCGTGGGT 660
Db 824 GAGTATTTAGTATTTTGAATATCATGATGATTCCTGATTAACGCTTCTGCGCGTGGGT 883
Qy 884 GTGGCTCAAGCACTCGGCAATTTGACGAAGAACGATTTAAGGAATGCGATGCGCG 943
Db 944 GTGGCTCAAGCACTCGGCAATTTGACGAAGAACGATTTAAGGAATGCGATGCGCG 1003
Qy 944 CCAGATCCGGAGCAATGAGAAATTTCCGCTGATAGTCGAGCGACGCTGGGCACTTT 1003
Db 1004 GTTAAATGGGTTGGCCGCAACGACGCTTCTTCACTTGAATATGAGAAACGTTTAA 1063
Qy 1064 GAAATCGGTTACCCGACCGATGATCCGATCATATGAACTGCGCGCAGACCGGTGC 900
Db 901 GATCGGACACAGCAATTCGCAAAATGACGTAATTCGCTTATATTTAGCAAGTGAATC 960
Qy 961 TTTAATCGGTGAACAACGAAACCGATGCTAAAGCCATGAAGGCAAAATTTCCGCA 1020
Db 1184 TTTAATCGGTGAACAACGAAACCGATGCTAAAGCCATGAAGGCAAAATTTCCGCA 1243
Qy 1021 GACAAATCGCATACCTAGAGATATAGTCAACAGATGAATTAATGAAATTTAAAGAA 1080
Db 1244 GACAAATCGCATACCTAGAGATATAGTCAACAGATGAATTAATGAAATTTAAAGAA 1303
Qy 1081 AGAATGACAAACCGTGTCAATATATGCGTGGCAATATTCATGATGCGCAGACCTTTA 1140
Db 1304 AGAATGACAAACCGTGTCAATATATGCGTGGCAATATTCATGATGCGCAGACCTTTA 1363
Qy 1141 ATTGAATAATCCAGGAATCAAGSTAAAGAGTTCGTAAAGC 1182
Db 1364 ATTGAATAATCCAGGAATCAAGSTAAAGAGTTCGTAAAGC 1405

```

```

RESULT 5
AAf82255
ID AAf82255 standard; DNA; 1182 BP.
XX
AC AAf82255;
XX
DT 21-JUN-2001 (first entry)
XX
DE Bacillus subtilis ffo 336 DNA encoding a PGA synthesizing enzyme.
XX
XX Bacterial; poly-gamma-L-glutamic acid; L-PGA;
XX poly-gamma-D,L-glutamic acid; PGA; glutamate racemase; db.
OS Bacillus subtilis.
FH Key Location/Qualifiers
FT 1..1182
FT CDS /*tag= a
PN JP2001017182-A.
XX
PD 23-JAN-2001.
XX
PF 09-JUL-1999; 99JP-00196335.
XX
PR 09-JUL-1999; 99JP-00196335.
XX
PA (NAGS) NAGASE SANGYO KK.
XX
DR WPI; 2001-285408/30.
DR P-PSDB; AAB74024.
XX
PT New nucleic acid encoding a glutamate racemase enzyme useful for the
PT preparation of poly-gamma-glutamic acid.
XX
PS Claim 2, Page 11-12; 17pp; Japanese.
XX
CC The present sequence encodes an enzyme which is useful in the production
CC of poly-gamma-L-glutamic acid (L-PGA) or poly-gamma-D,L-glutamic acid
CC (PGA). A plasmid comprising the present sequence may be used to transform
CC Escherichia coli. The transformants express the enzyme and PGA is
CC produced in the culture
XX
SQ Sequence 1182 BP; 399 A; 235 C; 272 G; 276 T; 0 U; 0 Other;
Query Match 99.1%; Score 1171.2; DB 4; Length 1182;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1173; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 7 TGGTATCATATATAGCCTGTGCTGTCTGATATGCTGATGCGAATATTAGAAAAACAGCA 66
Db 4 TGGTATCATATATAGCCTGTGCTGTCTGATATGCTGATGCGAATATTAGAAAAACAGCA 63
Qy 67 CATCAGAAAAACATTATATGCTCTCTCTGTTGGGTGAATATTAAAGGCAATCCGCGAATA 126
Db 64 CATCAGAAAAACATTATATGCTCTCTCTGTTGGGTGAATATTAAAGGCAATCCGCGAATA 123
Qy 127 TCGACTGTGCAAGGCTGACCAACCGGAATATTATAGAACCGGTTACAAAGCTGTTGA 186
Db 124 TCGACTGTGCAAGGCTGACCAACCGGAATATTATAGAACCGGTTACAAAGCTGTTGA 183
Qy 187 AAAACAACAGAAACAGATGCAAGATGATTACTGGGACACACCGGAGGAAAAAGCGATT 246
Db 184 AAAACAACAGAAACAGATGCAAGATGATTACTGGGACACACCGGAGGAAAAAGCGATT 243
Qy 247 AAAACGAACTTAAAGGCGCGAATATCGAGAGCAAAAAGAGTCATGAGAGAAACAGTA 306
Db 244 AAAACGAACTTAAAGGCGCGAATATCGAGAGCAAAAAGAGTCATGAGAGAAACAGTA 303
Qy 307 GAAAGAGGGGCTAACCGGATGTCAGTGAATGATGCTGTTAACCGAATTATCAATC 366

```


Db	502	CCTTATATATGGCCATCTTGTTCATTACAGATAGTGAATATATCCGAGTTCCTTTAAACAAAA	561
Qy	547	GCAAAAGACGAAACACAAAATGATCATATGTCGTATTAATCTCAAAAATTACAGATGAGTAT	606
Db	562	GCAAAAGACGAAACACAAAAGTCAATGTCGTATTAATCTCAAAAATTACAGATGAGTAT	621
Qy	607	TTTAGTAAATTTTGAATATCATGTATATCCCTGATATAGAGCTTCTCTGGCGCTGGGGTGGCT	666
Db	622	TTTAGTAAATTTTGAATATCATGTATATCCCTGATATAGAGCTTCTCTGGCGCTGGGGTGGCT	681
Qy	667	CAAGCACTCGGCATTGACGAAAGAACAGACATTTTAAGGAAATGCTGAATGCGCCGCAGAT	726
Db	682	CAAGCACTCGGCATTGACGAAAGAACAGACATTTTAAGGAAATGCTGAATGCGCCGCAGAT	741
Qy	727	CCGGAGACAAATAGAAATTTCTTCCTGATATCATGTATGATCCGAGCGAGCTGGGCACTTTGTTAAT	786
Db	742	CCGGAGACAAATAGAAATTTCTTCCTGATATCATGTATGATCCGAGCGAGCGCTGGGCACTTTGTTAAT	801
Qy	787	GGGTTTGGCGGAAACGACGCTTCTTCTTCACTTTGAATATATGAAACGATGTAAGAAATC	846
Db	802	GGGTTTGGCGGAAACGACGCTTCTTCTTCACTTTGAATATATGAAACGATGTAAGAAATC	861
Qy	847	GTTTACCAGACCGATGATCCGATCATCATGAACTGCGACGACCGTGTGATCGG	906
Db	862	GTTTACCAGACCGATGATCCGATCATCATGAACTGCGACGACCGTGTGATCGG	921
Qy	907	ACACGACAAATTCGCAATATGACGTATGCTTATATTTGAAGCAAGTGAATGATCTTAAATC	966
Db	922	ACACGACAAATTCGCAATATGACGTATGCTTATATTTGAAGCAAGTGAATGATCTTAAATC	981
Qy	967	GGTGAACCAACGAAACCGATGCTTAAGAACCTTATGAAGAGGCAAAATTCCTGACGACAA	1026
Db	982	GGTGAACCAACGAAACCGATGCTTAAGAACCTTATGAAGAGGCAAAATTCCTGACGACAA	1041
Qy	1027	CTGCATGACCTAGAGTATATGATCAACAGATGAAATTTATGAAATTTGTTAAAGAAAGATG	1086
Db	1042	CTGCATGATCTTAGAGTATATATCAACAGATGAAATTTATGAAATTTGTTAAAGAAAGATG	1101
Qy	1087	CACAAACCGTGCATATATATGCGCGTGGCAATATTCATGATGCGGACAGACCTTTAATTGAA	1146
Db	1102	CACAAACCGTGCATATATATGCGCGTGGCAATATTCATGATGCGGACAGACCTTTAATTGAA	1161
Qy	1147	AAATTCACGAATACAGGTAAAGCAGCTCGTAAC	1182
Db	1162	AAATTCACGAATACAGGTAAAGCAGCTCGTAAC	1197
RESULT 7			
ID	ABK75403	standard; DNA; 819 BP.	
XX	AC		
XX	XX		
XX	DT	13-AUG-2002 (first entry)	
XX	XX		
XX	DE	Bacillus licheniformis genomic sequence tag (GST) #2694.	
XX	KW	Differential gene expression; genomic sequenced tag; GST;	
KW	KM	altered culture condition; environmental stresses;	
KM	XX	physiological provocation; ds.	
OS	XX		
OS	XX	Bacillus licheniformis.	
PN	XX		
PN	XX	WO200229113-A2.	
PD	XX		
PD	XX	11-APR-2002.	
PF	XX		
PF	XX	05-OCT-2001; 2001WO-US031437.	
PR	XX		
PR	XX	06-OCT-2000; 2000US-00680598.	
XX	XX	27-MAR-2001; 2001US-0279526P.	
XX	XX		
PA		(NOVO) NOVOZYMES BIOTECH INC.	

	(NOVO) NOV02XYMES AS.
XA	Berka R., Clausen IG;
PI	WPI; 2002-416684/44.
XX	
PT	Monitoring differential expression of several genes in first Bacillus
PT	cell relative to expression of same genes in one or more second Bacillus
PT	cells, by using substrate containing Bacillus genomic sequenced tag
PT	array.
PS	
XS	Claim 4; SEQ ID NO 2694; 200pg.; English.
CC	The invention describes a method of monitoring differential expression of
CC	genes in a first Bacillus cell relative to expression of the genes in
CC	other Bacillus cells, comprising hybridising labelled nucleic acid probes
CC	isolated from Bacillus cells to a substrate containing array of Bacillus
CC	genomic sequenced tags (GST), examining the array, and determining
CC	a relative gene expression by an observed hybridisation reporter signal of
CC	a spot in the array. The method is useful for measuring the expression of
CC	genes in a first Bacillus cell relative to expression of the same genes
CC	in one or more second Bacillus cells. The method is useful for monitoring
CC	global expression of several genes from a Bacillus cell, discovering new
CC	genes, identifying possible functions of unknown open reading frames and
CC	monitoring gene copy number variation and stability. Monitoring changes
CC	in expression of genes may be used to provide a representation of the way
CC	in which Bacillus cells adapt to changes in culture conditions,
CC	environmental stress or other physiological provocation. Extensive follow
CC	-up characterisation is unnecessary, when one spot on an array equals one
CC	gene or one open reading frame, since sequence information is available.
CC	This sequence represents a genomic sequence tag (GST) used in the method
CC	of the invention. Note: The sequence data for this patent did not form
CC	part of the printed specification, but was obtained in electronic format
CC	directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
CC	
SX	Sequence 819 BP; 257 A; 166 C; 215 G; 180 T; 0 U; 1 Other:
QQ	
Query Match	46.4%; Score 548.4; DB 6; Length 819;
Best Local Similarity	79.9%; Pred. No. 2.3e-144;
Matches 645; Conservative 0; Mismatches 162; Indels 0; Gaps 0	
DQ	24 CTGTGCTGTCATACACTGTCATCGGAATTATAGAAAACAAGCAGCATCAAGAATAATTGA 83
DB	4 CTGTGTGATCGTGTGGGGGATCGGACATTATGAAAAAAAGGCCACCAGCAAATAATTCGA 63
OY	84 TGCCCCCTCCTGTGGCGTGGAATATTAACGGGCATCCGCGGAAAAATGACCTGTGACAAAGCT 143
DB	64 TCGGCTGCTTCCGAGTAGAACATBAACGGTTATACGGGAAGTCCAAGGTGACAAAGATT 123
OY	144 GACAACCCGAATATTATATAGAACCGCGTTACAGA CTGTGGAAAAACAAACGAAACAGA 203
DB	124 AACAAACAGGAAATTTATATCGAAGCAGGCTACAAAACAGTAGANNAACAAACCGGACAGA 183
OY	204 TGCAGAATGATTACTCTGGGACACACCCGAGAAAAAGCCGATTAAACGAAACCTCAGGG 263
DB	184 CGCAAGATGATTATTTTGGGACACACCCGGAAGAAGCCGATCAAAAAGAAACCCGMAAG 243
OY	264 GCCGAATATCGGAGAGCAAAAAGAAAGTCATGAGAGAAACAGTAGAAAAGAGGGCTTAACGC 323
DB	244 GCCGAATATTCGAGAGAGCAAGAGAGTTATGAAAGAAACGGTGGAAAGGGGCCAATGC 303
OY	324 GATTGTCA GTGATGATCATGGCTGTGTTAACCCAGATTATCAATCATCTTYCAGAGAAACT 383
DB	304 GATTGTCA GTGATGATGATGGCGCTTAATCTGTGATTACCAATCATCTTYCAGAGAAATT 363
OY	384 TTCTGACGCCAATATTCGGCGTCAATTGTGAAATGTTTTAGAAGCAANTAGATGTCATGGG 443
DB	364 GCTTCAGGCTAATATTCGGCGTATCGTGAACGTGTGGAGATCACATGATGATGAGG 423
OY	444 GGCGAGCTTGATGAAATTTGCAAGAGCGTTTACCGGTACAATTCCTTATATAGGCCATCT 503
DB	424 ACCGATTTTGATGAANAATCGCAAGACATTCACAGAACACATTCCTTATATAGGACATT 483

OY	504	GGTCAATTACAGTAGGATATATACGAGTTCCTTTAAACAAAAAGCAAAACGAAACAC	563
Db	484	GGTATTACTGATAGAGGATATACGATATTCCTTTAAACAAATGGCAAAAGAAAGAACAC	543
OY	564	AAAGATCATCATTTGGCTGATTAAGTCAAAAATTAACAGATGATATTTACGTAATTTGAATA	623
Db	544	AAAGATCATTCGTGCGACGACATTTCTAAATTAACAGATGATATTCAGACAGTTTAGTA	603
OY	624	CATGGTATTCCTCGATTAACGCTTCTCTGGCGCTGGGTGGCTCAAGCACTCGGCAATTGA	683
Db	604	CATGGTATTCCTCGATTAATAGCGCTCTCTTCGCTCGGTGATGCTCAAGCGTTGGGCATTGA	663
OY	684	CGAAGAAACAGATATTTAAGGAATGCTGAATGGCGCCAGATCCGGAGCAATGAAGAT	743
Db	664	CGAAGAAACCGCTCTTTAAAGGACGTGTAATGGCGCTCGATCCGGAGCCATVGGAAT	723
OY	744	TCTTCGCGTCAATCAGTCCGACGAGCGCTGGGCACTTTGTTAATGGGTTGGCCGCAACGA	803
Db	724	TCTGCGCGTCAATGAACGCCAAGATTCGCCGACATTCGTCAACGGTTTGGCGCAATGA	783
OY	804	CGCTTCTTCTACTTGAATATATGAA	830
Db	784	CGCAGCTTCACACTTTAAACATTTGAA	810
RESULT 8			
ID	ABN91998	standard; DNA; 1164 BP.	
AC	ABN91998;		
XX			
DT	24-JUL-2002	(first entry)	
DE			
XX			
KW	Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;		
KM	antibacterial; gene therapy; gene; ds.		
XX			
OS	Staphylococcus epidermidis.		
PN	US6380370-B1.		
XX			
PD	30-APR-2002.		
XX			
PF	13-AUG-1998; 98US-00134001.		
XX			
PR	14-AUG-1997; 97US-0055779P.		
PR	08-NOV-1997; 97US-0064964P.		
XX			
PA	(GENO-) GENOME THERAPEUTICS CORP.		
XX			
PI	Doucette-Stamm LA, Bush D;		
DR	WPI: 2002-381255/41.		
XX	P-PSDB; ABP39453.		
XX			
PT	Novel isolated nucleic acid encoding a Staphylococcus epidermis		
XX	polypeptide, useful for diagnosis and treating bacterial infections.		
PS	Disclosure; SEQ ID NO 1461; 267bp; English.		
XX			
CC	ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading		
CC	frame (ORF) nucleic acid sequences which encode the amino acid sequences		
CC	given in ABP35124 to ABP37960. The S. epidermidis sequences have		
CC	antibacterial activity and can be used in gene therapy. The sequences can		
CC	also be used in the diagnosis and treatment of bacterial infections,		
CC	particularly S. epidermidis infections. The sequences can be used to		
CC	screen for compounds able to interfere with the S. epidermidis life cycle		
CC	or inhibit S. epidermidis infection. N.B. The sequence data for this		
CC	percent did not form part of the printed specification, but was obtained		
CC	in electronic format directly from the USPTO web site		
XX			
SEQ	Sequence 1164 BP; 415 A; 172 C; 244 G; 333 T; 0 U; 0 Other;		

Query Match	28.6%;	Score 338.6;	DB 6;	Length 1164;
Best Local Similarity	59.7%;	Pred. No. 6e-85;		
Matches 588;	Conservative	0;	Mismatches 394;	Indels 3;
				Gaps 1;

OY	5	GCCTGGTACCTCATTAATAGCCGTGCTG---TCATACCTGCTCATGCGAAATATTAGCAAAAAC	61
Db	5	GATTGTATCTCATTAATAGCCGTGTGGCTTATTTCTCTGCTAGGAATTTAAAGAAAGA	64
OY	62	GACGACATCGAAMAAAACATTGATGCCCTCCCTGTTCCGGTGAATATTAAACGGATCCGCG	121
Db	65	AGCGTCATGCAAAATCGACTTGAAAAAATTCATTACGTAATTAACTAAATGGAATTTCTGTG	124
OY	122	GAAAATGCACTGTGACAAAGCTGACAAACCGGAATATTATATGAAGCCGGTTTACAAACTG	181
Db	125	GGAATCTTACCATTAATCTCGATGCGCTTAACAATGTCTTACGTAAGAACCAATATGAGATTG	184
OY	182	TTGGAAAAACAACAGAAACAGATGACAGAAATGATTTTACTGGGACACACCGAGAAAAAG	241
Db	185	TTGGTAAGACACACAGAAACAGATGACGAATGCTCTATTTGGTTTACTGAAAAAGGAATTC	244
OY	242	CGATTAAACGMAAACCTCAGGGGCGCAATATCGAGAGCAAAAAGAACTGAGAGAA	301
Db	245	CAGTAATCAGAAAAACCAAGGCCCAACATTGGAGAAACAACGAGATATTATTGCTAAG	304
OY	302	CAGTAAGAAAGGGGCTTACGCCATTGTCACTGAATGCAATGGCTTTTAAACCAATTTTC	361
Db	305	TGGTTAAACAAAAAGCAAAATGCCCTAGTAATGAGTATGCTGTAAATCAGATTTATC	364
OY	362	AAATCATCTTTCAGAGAAAGACTTCTGCAGGCCAATATCGCGCTCATTTGTAATGTTTTAG	421
Db	365	AAATTTACCTTTCAAAATGATTTAGTAAGAAATATCGGTATATGTTATATGATGAG	424
OY	422	AAGACCATATGATGTCAATGGGGCCGACGCTTGTATGAATTGCAAGACGTTTACCGCTA	481
Db	425	AAGACCATATGATGTCTTAGACCGACACTTTAAAGATGTAGCCGACAACTTTTACTGCA	484
OY	482	CAATTCCTTAATATGGCCATTTGTCAATTACAGATAGATATATACGAGTTCTTTTAAAC	541
Db	485	CAATTCATATACGGGAAATTAAGTTGTAAGAAAGATACATATATCTAGTTCTTTGTGA	544
OY	542	AAAAAGCAAAAGAACAAACAAAGCATCATTTGCTGATTAACCAAAAATTAACAGATG	601
Db	545	AGGAAGCTTAAABACGTATATTCAGAACCTCATTTGTTATGATTAAGACGTCATACAGAT	604
OY	602	AGTATTTACGTAAATTTTGAAATACATGATGTATTCCTGATTAACGCTTCTCTGCGCTGGGTG	661
Db	605	CATATTTTACGGAAATTCGATTATTAGATTTTCCGATTAATGATGCTATTGTTAGGAA	664
OY	662	TGGCTCAAGCACTGGGCAATTGACGAAGAAACAGCACTTAAAGGGAATGCTGATGCGCGC	721
Db	665	TAGGCAAGCACTTGGTGTATGATGAAGAACTGGCTTTCAAGAGTATGTTAAATGACCAAG	724
OY	722	CAGATCCGGGAGCAATGAAATTTCTCCGCTGATCAAGTCGAGCGAGCCTGGGCACTTTG	781
Db	725	CCGATCCAGGTGCTGTATGATTTAAATATTTCATGCAAAATCGCACAAAAATGATTTTG	784
OY	782	TTAATGGGTTTGGCCGAACGACGCTTCTTTACTTTGAATATATATGAAACGCTGTAAAG	841
Db	785	TTAATGTCAATTCGCTGTATATGAACCGCAGCTTCAAAAAGCGAATTTTAAATTAAGTGAGAT	844
OY	842	AAATTCGGTTACCCGACCGATGATCCGATCATCATCATGACCTGCGCGCAGACCGCTGTCG	901
Db	845	CATATTAATTATCCATACGATTAAGAAATATATCAATTTCTCAATTTGCTTCAAGTAAAGGTTG	904
OY	902	ATCGGACACAGCAATTCGCAATATGACGTATTCCTTAATTATGAAGCAAGTAACTGATCT	961
Db	905	ATAGAACCAACTCTTTGTGATTACTTTTAAAGGAAAGTGATTAAGCATGTTCTCATTT	964
OY	962	TAATTCGTGAACAACAGAACCGAT	986
Db	965	GTACAGGAAAAAGTACCAAAATGCT	989

RESULT 9
 ADS01519
 ID ADS01519 standard; DNA; 1164 BP.
 XX
 AC ADS01519;
 DT 04-NOV-2004 (first entry)
 XX
 DE Staphylococcus epidermidis polynucleotide seqid 814.
 XX
 KW antibacterial; vaccine; antisense therapy; Staphylococcus epidermidis;
 KW recombinant expression vector; infection; computer readable medium;
 KW computer based system; gene; ds.
 XX
 OS Staphylococcus epidermidis.
 XX
 PN US2004147734-A1.
 PD 29-JUL-2004.
 PF 01-DEC-2003; 2003US-00724972.
 PR 08-NOV-1997; 97US-0064966A.
 PR 13-AUG-1998; 98US-00134001.
 PR 29-NOV-1999; 99US-00450969.
 XX
 PA (DOUC/) DOUCETTE-STAMM L.
 PA (BUSH/) BUSH D.
 PI Doucette-Stamm L, Bush D;
 DR WPI; 2004-580138/56.
 DR P-Psdb; ADS05291.
 XX
 PT New isolated polypeptide, and encoding nucleic acid derived from
 PT Staphylococcus epidermidis, useful for diagnosing, preventing and/or
 PT treating an S. epidermidis bacterial infection.
 XX
 PS Claim 5, SEQ ID NO 814; 741bp; English.
 CC The invention describes an isolated nucleic acid comprising a nucleotide
 CC sequence with any of 3772 fully defined nucleotide sequences (SEQ ID NO:
 CC 1-3772) and encoding an Staphylococcus epidermidis polypeptide with any
 CC of 3772 fully defined amino acid sequences (SEQ ID NO: 3772-7544) as
 CC given in the specification. Also described are: a recombinant expression
 CC vector; a cell comprising a recombinant expression vector of (1);
 CC producing an S. epidermidis polypeptide; an isolated nucleic acid
 CC comprising a nucleotide sequence of at least 8 nucleotides in length; a
 CC vaccine composition for prevention or treatment of an S. epidermidis
 CC infection, comprising a nucleic acid cited above and a carrier; treating
 CC a subject for S. epidermidis infection; a recombinant or substantially
 CC pure preparation of an S. epidermidis polypeptide or its fragment; a
 CC vaccine composition for prevention or treatment of an S. epidermidis
 CC infection; detecting the presence of a Staphylococcus nucleic acid in a
 CC sample; a computer readable medium having recorded in it the nucleotide
 CC sequences with SEQ ID NO: 1-3772 or its fragments; a computer based
 CC system for identifying fragments of the Staphylococcus genome of
 CC commercial importance; a computer based system for identifying fragments
 CC of the Staphylococcus plasmids of commercial importance; identifying
 CC commercially important nucleic acid fragments of the Staphylococcus
 CC genome and/or plasmids; and identifying an expression modulating fragment
 CC of the Staphylococcus genome and/or plasmids. The methods and
 CC compositions of the present invention are useful for the diagnosis,
 CC prevention and/or treatment of an Staphylococcus epidermidis bacterial
 CC infection. This sequence encodes a S. epidermis protein of the invention.
 CC
 SQ Sequence 1164 BP; 415 A; 172 C; 244 G; 333 T; 0 U; 0 Other;

Query Match 28 6%; Score 338.6; DB 13; Length 1164;
 Best Local Similarity 59.7%; Pred. No. 66-85;
 Matches 588; Conservative 0; Mismatches 394; Indels 3; Gaps 1;

OY	5	GCTGTTACTCATATATAGCTGCTGCTG---TCATACTGCTCATCGAATATATAGAAAAAC	61
DB	5	GATTTGTTACTCATATATAGCTGCTGCTGCTGCTATCTCTGCTGAGATTAAGAGACAGA	64
OY	62	GAGCAATCAAAAAACATTGATGCTCCCTCTGTTGGGGTGAATATTAACGCATCCGCC	121
DB	65	AGCGTCATGCAAAATCGACTGAAAAAAATTCATTACGATTAATCAATAAAGAAATTCGTG	124
OY	122	GAAATGCACTGTCAGCAAGGCTGCAACCGGAATATTAATAGAACCGGTTTACAAGACTG	181
DB	125	GGAATCTTACGATTAATCTGAGTGGCTTACAGTGTCTTACGTCAGGACCATATAGACTTG	184
OY	182	TTGAAAAAACAACAGAACGATGCAAGATGATTTATCTGGACACACCGAGGAAAAAC	241
DB	185	TTGGTTAAGACACAGAAACAGATGCAAGATGCTTATGGTTTACGAAAAAGAAATATC	244
OY	242	CGATTAACGGAACCTTCAGGGGCGGAATATCGAGAGCAAAAAAGAGTCTAGAGAAA	301
DB	245	CAGTAAATCAGAAAACCAACAGGCGCCCAACATTTGAGAAACACAGAAATATTTGTAAGG	304
OY	302	CAGTAAAGAGGGGCTTACCGCATTTGTCAGTGAATGATGGCTTTAACCATTTATC	361
DB	305	TGGTTAAACAAAAAGCAAAATGCTTAAATGATGATGCTGTAATCAAGATTATC	364
OY	362	AAATCATCTTTCAGAAAGAACTTTCAGAGCCCAATATCGGCTGATTTGTAATGTTTAA	421
DB	365	AAATTTACTTTCAAAATGATTTAGTAAAGCAATATCGGTTAATGTTAATGATGATG	424
OY	422	AAGACCATATGATGTCATGAGGGGCGGACGCTTGATGAATGCAAGAGCTTTACCGCTA	481
DB	425	AAGACCATATGATGATGCTTACGACGCACTTAAAGATGAGCCAGACCTTTTACTGCA	484
OY	482	CAATTCCTTATTAATGAGCCATCTTTCATTAACATGATGATATATACGAGTCTTTAAAC	541
DB	485	CAATTCCTTATTAATGAGCCAAATTAATGATGATGATGATGATGATGATGATGATGATG	544
OY	542	AAAAAGCAAAAGAAAGCAAAACAAAGATCATGCTGATTAATCAAAAAATTAACAGATG	601
DB	545	AGGAAGCTTAAAGAGCTTAATTCAGAACTCATTTGTTGATGAATTAAGCCTTACAGAT	604
OY	602	AGTATTTACGTAATTTTGAATATACATGATGATTCCTGATTAACGCTTCTGCGGCTGGTG	661
DB	605	CATATTTACGTAATTTTGAATATTTAGTATTTCTGATTAATGATGATGATGATGATGATG	664
OY	662	TGGCTCAAGCACTCGGCAATTCAGAAAGAAACAGCATTTTAAGGAATGCTGAATGGCCGC	721
DB	665	TAGCCCAAGCAAGTGTGTGATGATGAAGAACTGCAATTAAGATGTTTAATGACACAG	724
OY	722	CAGATCCGGAGCAATGGAATTTCTCCGCTGATCAGTCCGAGCGAGCCTGGGCACTTTG	781
DB	725	CCGATCCAGGTCGTGTGATGAATTTAATTTTCCATGCAAAATGCAAAAAATGATTTTG	784
OY	782	TTAATGGGTTTCCCGCAACGACGCTTCTTCACTTTGAATATATGAAAACTGTAAAG	841
DB	785	TTAATGCAATTCGCTGCTGATTAAGAACGAGTCAAAAGCAATTTTAATTAAGTGAAT	844
OY	842	AAATGGTTATCCGACCGATGATCCGATCATATATGAATCAATCCGCGACGCTGTG	901
DB	845	CATATTAATTAATCATATGATGATTAAGAAATATATATCTCATTTGTGTGATGAGGTTG	904
OY	902	ATCGGACACAGCAATTCGAAATGACGATTTGCTTATTAATGAACGAAGTGAATGATCT	961
DB	905	ATGAGAACAAACCTCTTTGTTGTAATCTTTTATGATGATGATGATGATGATGATGATG	964
OY	962	TATCGTGAACACAGAACCGAT 986	
DB	965	GTAACGAAAAAGTACACAAATGT 989	

RESULT 10
 AAH54379
 ID AAH54379 standard; DNA; 3240 BP.

AAH54379;
 03-SEP-2001 (first entry)
 S. epidermidis genomic polynucleotide sequence SEQ ID NO:3743.
 Staphylococcus epidermidis SRI strain; infection; diagnosis; vaccination;
 endocarditis; ds.
 Staphylococcus epidermidis.
 MO200134809-A2.
 17-MAY-2001.
 09-NOV-2000; 2000WO-US030782.
 09-NOV-1999; 99US-0164258P.
 (GLAXO) GLAXO GROUP LTD.
 Kimmerly WJ;
 MPI; 2001-316495/33.
 Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
 useful for vaccinating against infections, e.g. endocarditis.
 Claim 8; Page 1366-1367; 2188pp; English.
 AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
 (II), given in AM81454 to AM81320, from Staphylococcus epidermidis. (I)
 and (II) can have antibacterial activity and therefore can be used in
 vaccination. The nucleic acids (I) may be used to produce the S.
 epidermidis polypeptides (II) via the production of vectors containing
 them which are used to produce hosts cells which express the
 polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
 used to vaccinate subjects and to raise antibodies against the bacteria.
 The polypeptides may also be used to assay for other inhibitors of their
 activity and therefore identify compounds that may be used for the
 treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
 AAH55090 represent specifically claimed S. epidermidis genomic DNA
 polynucleotide sequences from the present invention. AAH55091 to AAH55098
 represent oligonucleotide sequences and primers which are used in the
 exemplification of the present invention. N.B. The present invention
 specifically claims all the polynucleotide sequences given in the
 sequence listing of the present specification, however the sequence
 listing only goes up to SEQ ID NO:4454 so even though sequences are given
 in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present
 for SEQ ID NO:4455 to 4464
 Sequence 3240 BP; 1128 A; 464 C; 589 G; 1059 T; 0 U; 0 Other;
 Query Match 28.6%; Score 338.6; DB 4; Length 3240;
 Best Local Similarity 59.7%; Pred. NO. 9.5e-85;
 Matches 588; Conservative 0; Mismatches 394; Indels 3; Gaps 1;

CGATTAAACGGAACCTCAGGGGCCGAATATCGAGACAAAAAGAGTCATGAGAAA 301
 CAGTAATCAGAAAAACCAAGGGGCCCAACATTTGAGAGAACAGGATATTATTCGTAAGG 1972
 CAGTAGAAGAGGGGCTTAAGCCGATTTGAGTGAATGATGCTGTAAACCCAGATTATC 361
 TGGTTAAACAAAAAGCAATGCCCCCTAGTTAATGATGATGAGCTGTAATCCAGATTATC 2032
 AAATCATCTTTCAGGAAGAACTTCGAGGCCAATATCGCGCTCATGTGTAATTTTAA 421
 AAATTACCTTTCAAAATGATTTAGTAAAGCAATATCGGTGTAATTTGTAATGATG 2032
 AAGACATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 481
 AAGACATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2152
 CAATTCCTTATATGCGCATCTTGTCTATACAGATAGTGAATATCCGAGTCTTTAAAC 541
 CAATTCATATTAACGGAAATTAATGTTGTAATGAAGTAATCTAATAGTATTTTTC 2212
 AAAAGCAAAAGAACGAAACCAAAAGTCATCTGCTGTAATCTCAAAATATACAGATG 601
 AGAAGCTAAAGAACGTAATTCAGACTCATGTTGTAATGAAGACGTCATACAGAT 2272
 AGTATTTACGTAATTTTGAATACATGATGATGATGATGATGATGATGATGATGATG 661
 CATATTTACGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2332
 TGGCTCAGACATCGGCATTTGACGAGAAACAGCATTTAAGGAATCTGATGCGCCG 721
 TAGGCAAGCAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2392
 CAGATCGGAGCAATGAGATTTCCGCTGATCGAGCGAGCGCGGACATTTG 781
 CGATTCAGGCTGCTTGAATTAATTTTCCATGCAATTCGCAAAAAATGATTTG 2452
 TTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2512
 TTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 841
 AAATCGGTTACCGCAGCAGATGATGATGATGATGATGATGATGATGATGATGATG 901
 CATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2572
 ATGGAACAGCAATGCGAAATGATGATGATGATGATGATGATGATGATGATGATGAT 961
 ATGAACCAACCTTTGTTGATTAATTTTGAATGATGATGATGATGATGATGATGAT 2632
 TAATCGGTAACCAACAGAACCGAT 986
 GTACAGAAAAGTACCAAAATGAT 2657
 RESULT 11
 IDB06113
 ADB06113 standard; DNA; 1212 BP.
 ADB06113;
 20-NOV-2003 (first entry)
 Alloisococcus otilis antigenic protein encoding DNA SEQ ID NO:53.
 Alloisococcus otilidis; antigenic protein; immunogenic; immunisation;
 gene therapy; Gram-positive Bacterium; infection; gene; ds.
 Alloisococcus otilis.
 WO2003048304-A2.
 12-JUN-2003.
 25-NOV-2002; 2002WO-US036123.

MP	ADBI2064_11	1100001	1210000
MP	ADBI2064_12	1200001	1310000
MP	ADBI2064_13	1300001	1410000
MP	ADBI2064_14	1400001	1510000
MP	ADBI2064_15	1500001	1610000
MP	ADBI2064_16	1600001	1710000
MP	ADBI2064_17	1700001	1754382
ID	ADBI2064_standard	DNA; 1754382 BP.	
XX	ADBI2064;		
AC			
XX			
DT	20-NOV-2003	(first entry)	
DE	Alloiococcus otitis	entire genome sequence	SEQ ID NO:6651.
XX			
XX	Alloiococcus otitidis; antigenic protein; immunogenic; immunisation;		
KW	gene therapy; Gram-positive bacterium; infection; gene; de.		
XX			
OS	Alloiococcus otitis.		
XX			
PN	WO2003048304-A2.		
XX			
PD	12-JUN-2003.		
XX			
PF	25-NOV-2002; 2002WO-US036123.		
XX			
PR	29-NOV-2001; 2001US-033777P.		
PR	18-NOV-2002; 2002US-0426742P.		
XX			
PA	(NAHP) WYETH HOLDINGS CORP.		
XX			
F1	Fletcher LD, Mcmichael JC, Russell DP, Zagursky RJ;		
DR	WPI; 2003-505284/47.		
XX			
PT	New Alloiococcus otitidis polynucleotides and polypeptides, useful for		
PT	treating and diagnosing diseases, drug screening assays and monitoring of		
PT	effects during drug clinical trials.		
XX			
PS	Example 3; SEQ ID NO 6651; 1019P; English.		
XX			
CC	The present invention describes an isolated polynucleotide (1) of		
CC	Alloiococcus otitidis genomic DNA, which encodes an antigenic protein.		
CC	Alloiococcus otitidis is a Gram-positive bacterium. Also described: (1)		
CC	an isolated polypeptide that is encoded by the polynucleotide (1); (2) an		
CC	expression vector comprising the novel isolated polynucleotide (1), its		
CC	complement, degenerate variant or fragment; (3) a genetically engineered		
CC	host cell, transfected, transformed or infected with the vector of (2);		
CC	(4) an antibody specific for the polypeptide of (1); (5) an immunogenic		
CC	composition comprising the polypeptide, its complement, biological		
CC	equivalent or fragment, or the polynucleotide that is comprised in the		
CC	expression vector; (6) a pharmaceutical composition comprising the		
CC	polypeptide of (1) and a carrier; (7) a protein chip comprising an array		
CC	of the polypeptides of (1), their biological equivalent or fragment; (8)		
CC	immunising against Alloiococcus otitidis by administering to a host the		
CC	immunogenic composition; (9) detecting and/or identifying Alloiococcus		
CC	otitidis in the biological sample; (10) a kit comprising a container		
CC	containing the novel polynucleotide, its degenerate variant or fragment,		
CC	or the antibody of (4); and (11) producing a polypeptide by culturing the		
CC	genetically engineered host cell under conditions suitable to produce the		
CC	polypeptide from the culture. (1) can be used in gene therapy. The		
CC	polynucleotides, polypeptides, antibodies and compositions of the present		
CC	invention can be used for treating and diagnosing diseases, drug		
CC	screening assays and monitoring of effects during drug clinical trials.		
CC	The polynucleotides are useful for expressing and detecting Alloiococcus		
CC	otitidis. The present sequence represents the entire genome of		
CC	Alloiococcus otitidis, which is given in the exemplification of the		
CC	present invention.		
XX			
XX			
SQ	Sequence 1754382 BP; 484756A; 391265C; 387369G; 490992T; 0U; 00other;		
Query Match	26.5%;	Score 313.2;	DB 9; Length 110000;
Best Local Similarity	55.1%;	Pred No. 7.1e-77;	

	Matches	63%;	Conservative	0;	Mismatches	51%;	Indels	3;	Gaps	1;
QY	27	TCGTCGATCACTG	GTGATCATCGGAATATTTAGAAAAAGCAGCACTCA	GGAAAAACATTGATGC	86					
Db	23405	TGGTCCTTCCTA	TATGGCGCTTAGGGGTATTAAGAAACGACGCTTA	CGAGAAAAATATTA	23346					
QY	87	CCGCCCGTTCGG	GTGAATATTAACGGCATCCGCGGAAAAATGAC	CTGTGACAAAGGCTGAC	146					
Db	23345	AATCGACCTTAC	GGCATTATTAATTTATATGCAACGAGGCAACAAACG	TACTCTGTTTGT	23286					
QY	147	AACCGGAATATTA	TAATGAAGACCGGTTTACAAGCTGTGGAAAAACA	CAAGAAACAGATGC	206					
Db	23285	TACGGGGGGTCT	CAACAGACGCTGGCTTACCAAGTATTTGAAAAACA	CAAGAACTGACGTC	23226					
QY	207	AAGATATGATTT	ACTGGGACACACCGGAGAAAAAGCCGATTAA	CGGAAACCTCAGGGGCC	266					
Db	23225	ACCCATTATTTAC	TGTGGACCAAGCCAAAGAGGAGATCGAAAGGCTCC	ATCTTGAGCC	23166					
QY	267	GAATATCGAGAC	CAAAAAGAGCTATGAGAAACAGTAAAGAGGGGGCTTA	AGCGCAT	326					
Db	23165	TAATATTTCTGA	ACAAATACGTTATTTAAAAAGCAAAAAAGGGGGCTGAT	GTCCCT	23106					
QY	327	TGTCAGTGAATG	CATGGCTGTTAACCCAGATTATCAAAATCATCTTT	CAGAGAACTTCT	386					
Db	23105	TGATATTAAGAT	GTATGACAGTTATATCAAGATACCAAGATGTTTAT	TTCTTAATGAATGTT	23046					
QY	387	GCAAGGCCATAT	TCGGCGTCACTTGTGAATGTTTAAAGACCAATATGAT	GTATGAGGCC	446					
Db	23045	TCATCAAAATAT	TACAGCATCGTTAACGCTTTTGAAAGCCATATG	ATATCTTGATGCC	22986					
QY	447	GACGCTTGAATG	AAATTCGAAAGCGTTTACCCCTCAATTCCTTATAT	GGCCATCTGT	506					
Db	22985	AACAAATGATCA	AAATGGCTCGGGGCTATGCAAGACATCCCAAAATAT	GGGCTTGTTGT	22926					
QY	507	CATTACAGATAG	TAATATACGAGTCTTTTAAACAAAAGCAAAAAGAA	CGAAACAAA	566					
Db	22925	CATCTCAGATG	ATGACTATTAAGACTTTTATTTAGAGTACGCCAG	AGAGAAATTCGCCG	22866					
QY	567	AGTCATCATGTC	GATPACTCAAAAATTAACAGATGATTTTACGTA	TTTGAATACAT	626					
Db	22865	GGTGAAGGTAT	TAGAGCAACACAGGCGTTTAGTCAGAAATTA	TGATGATTCGCTTGTACATCT	22806					
QY	627	GGTATTCCTCG	ATAAACGCTTCTCTGGCGCTGGGTGTGCTCAAGCA	CTCGCATTTGACGA	686					
Db	22805	TGTCCTTCCCAA	CAATCTAGCTGTTCGCGTGTCTATTTGCTGAAGAA	CTCGACGTTGACCG	22746					
QY	687	AGAAACAGCAT	TTTAAAGGAATGCTGAATCCGCGCAGATCCG	GGAGCAATGAGATTC	746					
Db	22745	CGATACGTCAT	TAAAGATGTTTCAACTGACCTGATCCGGGCA	TATATGCAATTTGTT	22686					
QY	747	TCCGCTGATCA	GTGCGAGCGAGCTTGGGCACTTT--GTTAA	TGGGTTTGGCGAAACGA	803					
Db	22685	TGAGATAGGGGA	CCCTATTGACAGAGAGGTATTTCTAGCTTAT	GTGCTTGTCTGTATGTA	22626					
QY	804	CGCTTCTTTCT	ACTTTGAATTAATGAAAAGGTGAAAAGAAATTCG	GTATACCGACCGATGA	863					
Db	22625	ACCCACTTCTAC	ATCATGCTGATCATGAAAGCTCAAGAGAGGTGG	CTTAATAGTGAAG	22566					
QY	864	TCCGATCATCAT	GATGAACTGACCGCGCAGACCGGTGCAATCG	AGACACAGCAATTCGCAA	923					
Db	22565	CCCTATTTATTT	GAATCAATTTGCCGAGAAACCGCAGATGAC	CGGACTMAATTTATTCACCA	22506					
QY	924	TGACGTATTCG	CTTATATTGAAGCAAGTGAATCTTATTCGT	TGAACAACAGAAC	983					
Db	22505	AGACTTATTTG	GTCAAAATTAAGATGCCAGGGGTATTCG	CAATTTGGA	22446					
QY	984	GATCGTAAAA	CCCTATGAGAGGCAAAATTCCTGCAACAACTG	ATGACCTTGAAGTA	1043					
Db	22445	TATTTAGAAAG	GTATATAAAATCAAGTTCAGAGTCAGGGATTA	TTTATCTTTGGAAAG	22388					
QY	1044	TAACTCAACAT	GAAGAAATTAATGAAATTTGTTTAAAGAA	AAATATGCAACACG	1103					
Db	22385	GAATCTCTGGA	AGATATTTTACATGACCCCTTAATGACTTA	TGCTTCTGTGAAGTGTCTTT	22326					

QY 1104 TGGCGTCGCAATATTCATGTCGCCGACAGCCTTTAATGAAAAATCCAGAAATACA 1163
DB 22325 GGGTAATGGAATATTCATGTCGTGGAAGATTTTTCCTTACCTTGATGAAAAATTA 22266
QY 1164 GGTAAAGCAG 1173
DB 22265 CTTGTGGCAG 22256

RESULT 13

ADRA48536
ID ADRA48536 standard; DNA; 560 BP.

AC ADRA48536;
DT 04-NOV-2004 (first entry)
XX capB target sequence #3.
DE
XX Bacillus anthracis; cutaneous; respiratory anthrax infection; pagA; capB;
KW pXO1; pXO2; target sequence; ds.
XX
OS Bacillus anthracis.
PN MO2004070001-A2.
XX
PD 19-AUG-2004.
XX
PF 12-NOV-2003; 2003WO-US036240.
XX
PR 15-NOV-2002; 2002US-0426552P.
PR 16-MAY-2003; 2003US-0471082P.

PA (GENP-) GEN-PROBE INC.

PI Norman SA, Bungo JU, Hogan JU, Weisburg WG;

DR WPI; 2004-604428/58.

PT New oligonucleotides that hybridize specifically to a Bacillus anthracis
PT sequence, useful for detecting cutaneous and respiratory Bacillus anthrax
PT infections.

PS Claim 1; SEQ ID NO 34; 61bp; English.

XX The present invention relates to an oligonucleotide of 20-40 nucleotides
CC that specifically hybridizes to a sequence contained in a Bacillus
CC anthracis target sequence. The methods and compositions of the present
CC invention are useful for detecting the presence of Bacillus anthracis
CC nucleic acid in a sample, in particular for detecting cutaneous and
CC respiratory anthrax infections. Two synthetic genetic target sequences,
CC derived from pagA and capB gene sequences, were synthesized to provide
CC known standards for testing oligonucleotides for detection of the genes
CC carried by the plasmids pXO1 and pXO2, without requiring handling of
CC virulent Bacillus anthracis. The present sequence represents a capB
CC target sequence.

XX Sequence 560 BP; 181 A; 89 C; 131 G; 159 T; 0 U; 0 Other;

QY Query Match 24.9%; Score 294.2; DB 13; Length 560;

XX Best Local Similarity 70.6%; Pred. No. 1.6e-72;
Matches 392; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

DB 190 ACAACGAGAACAGATGCAAGATGATTTACTGGAACACCGAGAAACCGATTAA 249
XX
XX 1 ACAACGAGTAACATCTCGGAATGATATATGTTTACTGACGAGACCAACCGATTAA 60
QY 250 CGGAACCTCAGGGGCGGAATATCGAGAGCAAAAGAGTCATGAGAGAAACAGTAGAA 309
DB 61 CCGCCGTAAGAGAGTCTTATATCGGTGACCAACGAGGTACTTAAAGAGGCTGCTGAT 120
QY 310 AGAGGGGCTTAACGCGATTGTCTAGTGAATGATGAGCTGTTAAACCAAGATTATCAATATC 369

DB 121 TTAAAGAGAGAGCACTTATTTGTAATGATGAGCAGTTCAACCCGATTTATCAATATATC 180
QY 370 TTTCAGAGAGAACTTTCGACAGCCCAATATCGGCTCATTTGATGTTTGAAGACCAT 429
DB 181 TTCCAAATATAAATGATTTCAAGCAAAATGTTGAGTGTATTAATGTTTGAAGATCAT 240
QY 430 ATGATGTCATGGGGCGGACGCTTGATGAATTCAGAGAGGTTTACCGCTACAAATTCCT 489
DB 241 ATGATGTTTANGGACCTTACCTTGACAGAGTACAGTACAGCTTTCACCTGACATTTCA 300
QY 490 TATATGCGCATCTTGTCATTTACAGATGATGATATACCGAGTTCTTTAAACAAAGCA 549
DB 301 TATATGAGCATTTAGTACATGATGAAAGTAAATATCTTGATTTCTTAAAGAGTTTCA 360
QY 550 AAAGAGCAAAACAAAGATCATCTTCCTGATTAACCAAAATTAAGATGAGATTTA 609
DB 361 GAAGAGCAAAATTAACAAAGATGATTTGCGGATTAATTTACAAATTTCAAGAAATTTCTTA 420
QY 610 CGTATTTTGAATACATGATGATTCCTGATTAACGCTTCTGCGCTGGTGGCTCA 669
DB 421 CGAAATTTGATTTACATGCTCTTCCAGATTAATGATGATGCTTTAGCGGTAGCAAG 480
QY 670 GCACTCGGCATTTGACGAGAAACAGCATTTAAGGAAATGTAATGCGCCGAGATCG 729
DB 481 GCTCTGGGATTTGATGAGAAACAGCATTCGCTGTATGTAATGCTCATCGGATCA 540
QY 730 GGAGCAATGAGAAAT 744
DB 541 GGAGCAATGAGAAAT 555

RESULT 14

AAH53464
ID AAH53464 standard; DNA; 861 BP.

XX AAH53464;

XX 03-SEP-2001 (first entry)

DE S. epidermidis open reading frame nucleotide sequence SEQ ID NO:2321.

KW Staphylococcus epidermidis SRI strain; infection; diagnosis; vaccination;
KW endocarditis; ds.

OS Staphylococcus epidermidis.

PN MO200134809-A2.

PD 17-MAY-2001.

PF 09-NOV-2000; 2000WO-US030782.

PR 09-NOV-1999; 99US-0164258P.

PA (GLAXO) GLAXO GROUP LTD.

PI Kimmery WJ;

DR WPI: 2001-316495/33.

DR P-PSDB; AAG82614.

PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis.

PS Claim 8; Page 616-617; 2188bp; English.

XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis. (I)
CC and (II) can have antibacterial activity and therefore can be used in
CC vaccination. The nucleic acids (I) may be used to produce the S.
CC epidermidis polypeptides (II) via the production of vectors containing
CC them which are used to produce hosts cells which express the

CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to AAH55098
CC represent oligonucleotide sequences and primers which are used in the
CC exemplification of the present invention. N.B. The present invention
CC specifically claims all the polynucleotide sequences given in the
CC sequence listing of the present specification, however the sequence
CC listing only goes up to SEQ ID NO:4454 so even though sequences are given
CC in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present
CC for SEQ ID NO:4455 to 4464

XX Sequence 861 BP; 311 A; 123 C; 175 G; 252 T; 0 U; 0 Other;

Query Match 20.8%; Score 246.2; DB 4; Length 861;

Best Local Similarity 60.0%; Pred. No. 7.3e-59;

Matches 410; Conservative 0; Mismatches 273; Indels 0; Gaps 0;

```
QY 304 GTAGAAAAGGGGCTAAGCGATGTGTCAGTGAATGCAATGCGCTGTTAACCAGATTATCA 363
DB 4 GTTAAACAAAAGAGAAATGCCCTAGTTATAGATGCTGTAATCCAGATTATCA 63
QY 364 ATCATCTTTCAGAAAGAACTTCGACAGCCATAATCGCGCTCATTTGTAATGTTTAA 423
DB 64 ATTACCTTTCAAAATGATTAGTAAAGCAAAATCGGTATTTGTTAATGTATGAA 123
QY 424 GACCATATGATGTCATGAGGCGGACGCTTANTGAAATGGAAGCGCTTACCGCTACA 483
DB 124 GACCATATGATGTCATGAGGCGGACGCTTANTGAAATGGAAGCGCTTACCGCA 183
QY 484 ATTCCTTATATGGCCATCTTGCATTAACAATAGTGAATATACCGAATCTTTAAACA 543
DB 184 ATTCATATACGGGAATTTGTTGTAATGAAGATACTATACAGTTCTTTGCAAG 243
QY 544 AAACAAAAGAGCAAAACACAAAAGTCATCTGCTGATTAATCAAAATATACAGATAG 603
DB 244 GAACCTAAAAGCGTAAATTCAGAACTCTTTGTTGATTAAGACGTATACAGAAATCA 303
QY 604 TATTTACGTAATTTGTAATACATGATATCCCTGATTAACGCTTCTGCGCGCTGGGTG 663
DB 304 TATTTACGGAATTTGATTTATTTAGTATTTCTGATTAATGATGTTGTTAGAAATA 363
QY 664 GCTAAGACCTCGGATTCAGGAAGAAACACATTTTAAGGAATGCTGAATGCCCGCA 723
DB 364 GCGAAGAGATTGGTATGATGAAGAACTGCATTTACAAAGGTATGTTAAATGACACAG 423
QY 724 GATCCGGAGAGATGAGAAATCTTCGCTGATTCAGTCCGAGCGAGCCCTGGGCACTTGT 783
DB 424 GATCCAGGTGCTGTTAGATTAATTTCCATGCAAAATCCGCAAAAATGATTTGTT 483
QY 784 AATGGGTTTGGCGGAAAGCAGCGCTTCTTCTAATTTGTAATATATGAAACGTGTAAAG 843
DB 484 AATGCATTCGCTGCTAATGAACCGCAGCTCAAAAAGGATTTTAAATAGTGAATCA 543
QY 844 ATCGGTTAACCGGACGATGATCCGATCATCATGAACTGCGGCGAGACCGGTGAT 903
DB 544 TATAATTAATCATATGATTAAGAAATAATATATTCATTCATTTGCTTCGATAGGGTTGAT 603
QY 904 CGGACACAGCAATTCGCAATATGACATATGCTTATTTATGGAAGCAAGCACTGATCTTA 963
DB 604 AGAACACAACTCTTTGTTGATTAATCTTTTAGGTGAAGTCGATGCAATGTTCTCATTTGT 663
QY 964 ATCGGTGAACACAGAAACGAT 986
DB 664 ACAGAAAAAGTACAAATGAT 686
```

RESULT 15

ABK75876

_ID ABK75876 standard; DNA; 309 BP.

```
XX ABK75876;
AC 13-AUG-2002 (first entry)
DT Bacillus licheniformis genomic sequence tag (GST) #3167.
DE Differential gene expression; genomic sequenced tag; GST;
KW altered culture condition; environmental stress;
KM physiological provocation; de.
XX Bacillus licheniformis.
OS WO200229113-A2.
PN 11-APR-2002.
PD 05-OCT-2001; 2001WO-US031437.
PF 06-OCT-2000; 2000US-00680598.
PR 27-MAR-2001; 2001US-0279526P.
XX (NOVO ) NOVOZYMES BIOTECH INC.
PA (NOVO ) NOVOZYMES AS.
PI Berka R, Clausen IG;
P1 WPI: 2002-416684/44.
DR Monitoring differential expression of several genes in first Bacillus
PT cell relative to expression of same genes in one or more second Bacillus
PT cells, by using substrate containing Bacillus genomic sequenced tag
PT array.
XX Claim 4; SEQ ID NO 3167; 200pp; English.
XX The invention describes a method of monitoring differential expression of
XX genes in a first Bacillus cell relative to expression of the genes in
XX other Bacillus cells, comprising hybridising labelled nucleic acid probes
XX isolated from Bacillus cells to a substrate containing array of Bacillus
XX genomic sequenced tags (GST), examining the array, and determining
XX relative gene expression by an observed hybridisation reporter signal of
XX a spot in the array. The method is useful for measuring the expression of
XX genes in a first Bacillus cell relative to expression of the same genes
XX in one or more second Bacillus cells. The method is useful for monitoring
XX global expression of several genes from a Bacillus cell, discovering new
XX genes, identifying possible functions of unknown open reading frames and
XX monitoring gene copy number variation and stability. Monitoring changes
XX in expression of genes may be used to provide a representation of the way
XX in which Bacillus cells adapt to changes in culture conditions,
XX environmental stress or other physiological provocation. Extensive follow
XX up characterisation is unnecessary, when one spot on an array equals one
XX gene or one open reading frame, since sequence information is available.
XX This sequence represents a genomic sequence tag (GST) used in the method
XX of the invention. Note: The sequence data for this patent did not form
XX part of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX Sequence 309 BP; 100 A; 62 C; 76 G; 71 T; 0 U; 0 Other;
Query Match 15.3%; Score 181; DB 6; Length 309;
Best Local Similarity 74.1%; Pred. No. 1.3e-40;
Matches 229; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
```

```
QY 873 CATATGAACCTGCGGCGAGACCGTGTGATCGGACACAGCAATTCGCAATGACGAT 932
DB 1 CCTTATTTACTGCGCGCGCGACAGGTTAGACAGAAACAGCAATTTCCGGAAGTGTCT 60
QY 933 GCGTTATATGAAGCAAGTGAAGTATCTTAATGCGTGAAGCAACAGAACGATCGTAA 992
DB 61 TCCTTATATTTGAACCAAGTGAAGTGTGCTGATTTGAGAAACACAGAGCCGATGCTAA 120
QY 993 AGCCTATGAAGAAAGCAAAATTTCTGCGAGCAACAACTGCATGACCTAGATTAAGTCAAC 1052
```


Db 121 AGCATATGAGCAGGCAAAATTCTGCGACAAAGCTGTTGATTGAGCAAAATCAAC 180
QY 1053 AGATGAAATTATGAAATTGTTAAGAAAAAGAAATGCAAAACCGTGTATATATATGCGCTCGG 1112
Db 181 GGAAGAAATCATGTTCAATGCTGAAAAACAAGCTTGAGGGCCGCGTTATTACGAGTCGG 240
QY 1113 CAATATTCATGCTGCCGAGAGCCCTTTAATTGAAAAATCCAGAAATACAGGTAAAGCA 1172
Db 241 AAATATCCAGGAGCAGCGAGCTCTCATGTGAAAAAATGCAAGATTACAGATTAAAGCA 300
QY 1173 GTCGTTAG 1181
Db 301 GTTCGTTAG 309

Search completed: February 27, 2006, 07:05:15
Job time : 756.404 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:34:28 ; Search time 1294.32 Seconds
(without alignments)
7551.777 Million cell updates/sec

Title: US-10-789-164-1

Perfect score: 1182

Sequence: 1 atggagctggcttaccatcatc.....agtaagcagctcgtaagc 1182

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA_Main:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1182	100.0	1182	8	US-10-484-605-1
2	1182	100.0	1182	9	US-10-789-164-1
3	1182	100.0	6536	8	US-10-484-605-6
4	548.4	46.4	819	3	US-09-974-300-2694
5	338.6	28.6	1164	7	US-10-724-972A-814
6	313.2	26.5	1212	9	US-10-501-282-53
7	313.2	26.5	1754382	9	US-10-501-282-6651
8	181	15.3	309	3	US-09-974-300-3167
9	44.4	3.8	1338	7	US-10-282-122A-10762
10	42.6	3.6	600	7	US-10-021-323-10884
11	42.2	3.6	4254	3	US-09-814-353-17917
12	42	3.6	12507	6	US-10-311-455-272
13	41.4	3.5	3673778	6	US-10-312-841-1
14	41	3.5	6000	6	US-10-311-455-2278
15	40	3.4	7312	7	US-10-311-455-1788
16	40	3.4	7312	7	US-10-221-613-312
17	39.8	3.4	11944	6	US-10-311-455-2159
18	39.6	3.4	1047	6	US-10-724-972A-3451
19	39.6	3.4	1546	6	US-10-017-161-2325
20	39.6	3.4	1546	6	US-10-292-798-1971
21	39.2	3.3	519	5	US-10-027-632-68820
22	39.2	3.3	519	5	US-10-027-632-68821
23	39.2	3.3	519	5	US-10-027-632-68822

C	24	39.2	3.3	519	5	US-10-027-632-298134	Sequence 298134,
C	25	39.2	3.3	519	5	US-10-027-632-298135	Sequence 298135,
C	26	39.2	3.3	519	6	US-10-027-632-68820	Sequence 68820, A
C	27	39.2	3.3	519	6	US-10-027-632-68821	Sequence 68821, A
C	28	39.2	3.3	519	6	US-10-027-632-68822	Sequence 68822, A
C	29	39.2	3.3	519	6	US-10-027-632-298134	Sequence 298134,
C	30	39.2	3.3	519	6	US-10-027-632-298135	Sequence 298135,
C	31	39.2	3.3	519	6	US-10-027-632-298135	Sequence 298135,
C	32	38.8	3.3	7057	6	US-10-311-455-147	Sequence 147, App
C	33	38.8	3.3	7057	6	US-10-240-485-147	Sequence 147, App
C	34	38.8	3.3	7057	7	US-10-221-613-317	Sequence 317, App
C	35	38.8	3.3	7057	7	US-10-221-613-317	Sequence 317, App
C	36	38.6	3.3	1008	7	US-10-282-122A-34706	Sequence 34706, A
C	37	38.6	3.3	5311	6	US-10-311-455-992	Sequence 992, App
C	38	38.4	3.2	511	7	US-10-021-333-771	Sequence 771, App
C	39	38.4	3.2	1353	7	US-10-282-122A-20958	Sequence 20958, A
C	40	38.4	3.2	11597	8	US-09-070-927A-222	Sequence 222, App
C	41	38.4	3.2	219715	3	US-10-417-375-63	Sequence 63, Appl
C	42	38.2	3.2	934	6	US-10-164-758-3	Sequence 3, Appl
C	43	38.2	3.2	1131	7	US-10-282-122A-34649	Sequence 34649, A
C	44	38.2	3.2	1158	7	US-10-724-972A-3149	Sequence 3149, Ap
C	45	38	3.2	599	5	US-10-027-632-128367	Sequence 128367,

ALIGNMENTS

RESULT 1
US-10-484-605-1
Sequence 1, Application US/10484605
Publication No. US20040253704A1
GENERAL INFORMATION:
APPLICANT: Sung, Moon-Hee
APPLICANT: Hong, Seung-Pyo
APPLICANT: Lee, Jong-Su
APPLICANT: Jung, Chang-Min
APPLICANT: Kim, Chul-doong
APPLICANT: Soda, Kenji
APPLICANT: Ashiuchi, Makoto
TITLE OF INVENTION: SURFACE EXPRESSION VECTORS HAVING pSBCA, THE GENE CODING
TITLE OF INVENTION: POLY-GAMMA-GLUTAMATE SYNTHETASE, AND A METHOD FOR EXPRESSION OF
TARGET PROTEIN AT THE SURFACE OF MICROORGANISM USING THE VECTOR
FILE REFERENCE: 4240-101
CURRENT APPLICATION NUMBER: US/10/484,605
CURRENT FILING DATE: 2004-01-20
PRIOR APPLICATION NUMBER: PCT/KR02/01522
PRIOR FILING DATE: 2002-08-09
PRIOR APPLICATION NUMBER: KR 2001-48373
PRIOR FILING DATE: 2001-08-10
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn version 3.2
SEQ ID NO 1
LENGTH: 1182
TYPE: DNA
ORGANISM: Bacillus subtilis
US-10-484-605-1

Query Match 100.0%; Score 1182; DB 8; Length 1182;
Best Local Similarity 100.0%; Pred. No. 7.6e-312;
Matches 1182; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATGGGCTGGTACTCTATTAAGCTTGTCTGCATCTGTGATCGGAATATTAGAAAA	60
DB	1	ATGGGCTGGTACTCTATTAAGCTTGTCTGCATCTGTGATCGGAATATTAGAAAA	60
QY	61	CGAGCATCAGAAAAATTGATGCCCTCGCTGGGTGAATTTAACGGCATCCGC	120
DB	61	CGAGCATCAGAAAAATTGATGCCCTCGCTGGGTGAATTTAACGGCATCCGC	120
QY	121	GGAATTCGACTGTGCAAGGCTGACACCGGAATTTAATAGACCGGTTCAAGACT	180
DB	121	GGAATTCGACTGTGCAAGGCTGACACCGGAATTTAATAGACCGGTTCAAGACT	180

181 GTTGGAAAAACAAGAGAGATGCAAGATTTTCTGCGACACACCGGAGAAAAG 240
181 GTTGGAAAAACAAGAGAGATGCAAGATTTTCTGCGACACACCGGAGAAAAG 240
241 CCGATTAAACGGAACCTCAGGCGCGGAATATCGGAGGCAAAAAGAGTATGAGAA 300
241 CCGATTAAACGGAACCTCAGGCGCGGAATATCGGAGGCAAAAAGAGTATGAGAA 300
301 ACAGTAGAAAGAGGCGCTAAGCGCATTTGTCACTGTAATGCAAGGCTGTTACCCGATAT 360
301 ACAGTAGAAAGAGGCGCTAAGCGCATTTGTCACTGTAATGCAAGGCTGTTACCCGATAT 360
361 CAAATCATCTTTCAGAGAGAACTTCTGCAAGGCAATATCGGCGTCAATTTGTAATGTTTA 420
361 CAAATCATCTTTCAGAGAGAACTTCTGCAAGGCAATATCGGCGTCAATTTGTAATGTTTA 420
421 GAAGACCATATGATGTCATGGGCGGACGCTTGTAATGTAATGCAAGGCTTTACCGCT 480
421 GAAGACCATATGATGTCATGGGCGGACGCTTGTAATGTAATGCAAGGCTTTACCGCT 480
481 ACAATTCCTTAAATGCGCATCTTGTCACTAAGATAGTAATGCAAGGCTTTTAA 540
481 ACAATTCCTTAAATGCGCATCTTGTCACTAAGATAGTAATGCAAGGCTTTTAA 540
541 CAAAAAGCAAAAG 600
541 CAAAAAGCAAAAG 600
601 GAGTATTTACGTAATTTTGAATATCATGATATTCCTGTAATACGCTTCTGCGCGTGGGT 660
601 GAGTATTTACGTAATTTTGAATATCATGATATTCCTGTAATACGCTTCTGCGCGTGGGT 660
661 GTGGCTCAAGACCTCGGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
661 GTGGCTCAAGACCTCGGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
721 CCAGATCCGAG 780
721 CCAGATCCGAG 780
781 GTTAATGGGTTTGGCGCAACGAGCGCTTCTTCACTTTGAATATATGAAACGTTTAA 840
781 GTTAATGGGTTTGGCGCAACGAGCGCTTCTTCACTTTGAATATATGAAACGTTTAA 840
841 GAAATCGGTTTCCGAG 900
841 GAAATCGGTTTCCGAG 900
901 GATCGAGACAGCAATTCGCAATGAGAGATTCCTTATATTTGAAGAGAGAGAGAGAGAG 960
901 GATCGAGACAGCAATTCGCAATGAGAGATTCCTTATATTTGAAGAGAGAGAGAGAGAG 960
961 TTAATCGGTAAGCAACAG 1020
961 TTAATCGGTAAGCAACAG 1020
1021 GACCAATCGAG 1080
1021 GACCAATCGAG 1080
1081 AGAATGCAACAG 1140
1081 AGAATGCAACAG 1140
1141 ATTGAAAAAATCGAG 1182
1141 ATTGAAAAAATCGAG 1182

RESULT 2
US-10-789-164-1
; Sequence 1, Application US/10789164
; Publication No. US20050191720A1

GENERAL INFORMATION:
; APPLICANT: Sun, M.H. et al.
; TITLE OF INVENTION: SURFACE EXPRESSION METHOD OF PEPTIDES P5 AND ANAL3 USING THE GENE
; TITLE OF INVENTION: ENCODING POLY-GLUTAMATE SYNTHETASE
; FILE REFERENCE: P1574
; CURRENT APPLICATION NUMBER: US/10/789,164
; CURRENT FILING DATE: 2004-02-27
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: KopelecIn 1.71
; SEQ ID NO 1
; LENGTH: 1182
; TYPE: DNA
; ORGANISM: Bacillus subtilis
US-10-789-164-1

Query Match 100.0%; Score 1182; DB 9; Length 1182;
Best Local Similarity 100.0%; Pred. No. 7.6e-312; Indels 0; Gaps 0;
Matches 1182; Conservative 0; Mismatches 0;

1 ATGGGCTGTTACTCATTAATAGCCTGCTGTCACTGTCATGTCATGCAATATTTGAAAAA 60
1 ATGGGCTGTTACTCATTAATAGCCTGCTGTCACTGTCATGTCATGCAATATTTGAAAAA 60
61 CGACGACATCAAGAAACATTGATGCCCTCCTGTTGGGTGAATTTAACGGCATCCGC 120
61 CGACGACATCAAGAAACATTGATGCCCTCCTGTTGGGTGAATTTAACGGCATCCGC 120
121 GGAATATGACTGTGACAGAGGCTGACAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
121 GGAATATGACTGTGACAGAGGCTGACAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
181 GTTGGAAAAACAAG 240
181 GTTGGAAAAACAAG 240
241 CCGATTAAACGGAACCTCAGGCGCGGAATATCGGAGGCAAAAAGAGTATGAGAA 300
241 CCGATTAAACGGAACCTCAGGCGCGGAATATCGGAGGCAAAAAGAGTATGAGAA 300
301 ACAGTAGAAAGAGGCGCTAAGCGCATTTGTCACTGTAATGCAAGGCTTTTAA 360
301 ACAGTAGAAAGAGGCGCTAAGCGCATTTGTCACTGTAATGCAAGGCTTTTAA 360
361 CAAATCATCTTTCAGAGAGAACTTCTGCAAGGCAATATCGGCGTCAATTTGTAATGTTTA 420
361 CAAATCATCTTTCAGAGAGAACTTCTGCAAGGCAATATCGGCGTCAATTTGTAATGTTTA 420
421 GAAGACCATATGATGTCATGGGCGGACGCTTGTAATGTAATGCAAGGCTTTACCGCT 480
421 GAAGACCATATGATGTCATGGGCGGACGCTTGTAATGTAATGCAAGGCTTTACCGCT 480
481 ACAATTCCTTAAATGCGCATCTTGTCACTAAGATAGTAATGCAAGGCTTTTAA 540
481 ACAATTCCTTAAATGCGCATCTTGTCACTAAGATAGTAATGCAAGGCTTTTAA 540
541 CAAAAAGCAAAAG 600
541 CAAAAAGCAAAAG 600
601 GAGTATTTACGTAATTTTGAATATCATGATATTCCTGTAATACGCTTCTGCGCGTGGGT 660
601 GAGTATTTACGTAATTTTGAATATCATGATATTCCTGTAATACGCTTCTGCGCGTGGGT 660
661 GTGGCTCAAGACCTCGGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
661 GTGGCTCAAGACCTCGGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
721 CCAAGATCCGAG 780
721 CCAAGATCCGAG 780
781 GTTAATGGGTTTGGCGCAACGAGCGCTTCTTCACTTTGAATATATGAAACGTTTAA 840
781 GTTAATGGGTTTGGCGCAACGAGCGCTTCTTCACTTTGAATATATGAAACGTTTAA 840

Db 781 GTTATGGGTTTGGCCGCAACGACGCTTCTTACTTTGAATATATGAAACGTGTAAA 840
Qy 841 GAAATCGGTTACCGGACCGGATGATCCGATCATCATGAACTGCCGGCGACCGTGTTC 900
Db 841 GAAATCGGTTACCGGACCGGATGATCCGATCATCATGAACTGCCGGCGACCGTGTTC 900
Qy 901 GATCGGACACAGCAATTTGGCAATAACGATGTCCTTTATTTGAAAGCAAGTGAATCATC 960
Db 901 GATCGGACACAGCAATTTGGCAATAACGATGTCCTTTATTTGAAAGCAAGTGAATCATC 960
Qy 961 TTAATCGGTAACACAGCAACCGATCGTAAAGCCTTATGAAAGGCAAAATTCCTGA 1020
Db 961 TTAATCGGTAACACAGCAACCGATCGTAAAGCCTTATGAAAGGCAAAATTCCTGA 1020
Qy 1021 GACAAATCGCATGACCTAGATATTAAGTCAACAGATGAATAATTTGTAAGAAA 1080
Db 1021 GACAAATCGCATGACCTAGATATTAAGTCAACAGATGAATAATTTGTAAGAAA 1080
Qy 1081 AGAATGCAACACCGGTGTCTATATGCGGTGCGCAATTTTCATGTCGCCGACAGCTTTA 1140
Db 1081 AGAATGCAACACCGGTGTCTATATGCGGTGCGCAATTTTCATGTCGCCGACAGCTTTA 1140
Qy 1141 ATTGAAAAATCCAGAAATACAGGTAAGAGCTGTAGC 1182
Db 1141 ATTGAAAAATCCAGAAATACAGGTAAGAGCTGTAGC 1182

RESULT 3
US-10-484-605-6

/ Sequence 6, Application US/10484605
/ Publication No. US20040253704A1
/ GENERAL INFORMATION:
/ APPLICANT: Sung, Moon-Hee
/ APPLICANT: Hong, Seung-Pyo
/ APPLICANT: Lee, Jong-Su
/ APPLICANT: Jung, Chang-Min
/ APPLICANT: Kim, Chul-Joong
/ APPLICANT: Soda, Kenji
/ APPLICANT: Ashiuchi, Makoto
/ TITLE OF INVENTION: SURFACE EXPRESSION VECTORS HAVING PG8CA, THE GENE CODING
/ TITLE OF INVENTION: POLY-GAMMA-GLUTAMATE SYNTHETASE, AND A METHOD FOR EXPRESSION OF
/ TITLE OF INVENTION: TARGET PROTEIN AT THE SURFACE OF MICROORGANISM USING THE VECTOR
/ FILE REFERENCE: 4240-101
/ CURRENT APPLICATION NUMBER: US/10/484, 605
/ PRIOR FILING DATE: 2004-01-20
/ PRIOR APPLICATION NUMBER: PCT/KR02/01522
/ PRIOR FILING DATE: 2002-08-09
/ PRIOR APPLICATION NUMBER: KR 2001-48373
/ NUMBER OF SEQ ID NOS: 8
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 6
/ LENGTH: 6536
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Synthetic Construct
US-10-484-605-6

Query Match 100.0%; Score 1182; DB 8; Length 6536;
Best Local Similarity 100.0%; Pred. No. 1,9e-311;
Matches 1182; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGGGCGGTACTCATATAGCTGCTGTCTCATCTGTCATCGGATATTTAGAAAA 60
Db 224 ATGGGCGGTACTCATATAGCTGCTGTCTCATCTGTCATCGGATATTTAGAAAA 283
Qy 61 CGAGCATCAGAAAAACATTTGATCCCTCCCTGTTGGGTGAATATTAACGGCATCCG 120
Db 284 CGAGCATCAGAAAAACATTTGATCCCTCCCTGTTGGGTGAATATTAACGGCATCCG 343
Qy 121 GGAATCGACTGTGACAGGCTGACAAACCGAATATTTATGAAAGCGGTTACAGACT 180

Db 344 GGAATTCGACTGTGACAGGCTGACAAACCGAATATTTATGAAAGCGGTTACAGACT 403
Qy 181 GTTGAATAAACAAGGAACGATGCAAGATGATTTACTGGGACACACCGGAGAAAA 240
Db 404 GTTGAATAAACAAGGAACGATGCAAGATGATTTACTGGGACACACCGGAGAAAA 463
Qy 241 CCGATTAACGGAAACCTCAGGGGCGGAAATTCGAGAGCAAAAAAGATCATGAGAA 300
Db 464 CCGATTAACGGAAACCTCAGGGGCGGAAATTCGAGAGCAAAAAAGATCATGAGAA 523
Qy 301 ACAGTAGAAAGAGGGCTTACCGCATTTGTAGTATGATGATGCTGTTTACCCAGATAT 360
Db 524 ACAGTAGAAAGAGGGCTTACCGCATTTGTAGTATGATGATGCTGTTTACCCAGATAT 583
Qy 361 CAATCATCTTTCCAGAAACCTTCTGAGGCGCAATATCGGCGCATTTGTAATTTTA 420
Db 584 CAATCATCTTTCCAGAAACCTTCTGAGGCGCAATATCGGCGCATTTGTAATTTTA 643
Qy 421 GAAGACATATGATGTATGAGGCGGACGCTTGATGAAATTTGCAAGCGTTTACGCT 480
Db 644 GAAGACATATGATGTATGAGGCGGACGCTTGATGAAATTTGCAAGCGTTTACGCT 703
Qy 481 ACAATTCCTTATATATGATGATGCTTGTCTATTAAGATGATGATATCCGATTTTAA 540
Db 704 ACAATTCCTTATATATGATGATGCTTGTCTATTAAGATGATGATATCCGATTTTAA 763
Qy 541 CAAAAAGCAAAAGCAAAACAAAGTCAATCTGCTGATTAACCTCAAAATTAACAGT 600
Db 764 CAAAAAGCAAAAGCAAAACAAAGTCAATCTGCTGATTAACCTCAAAATTAACAGT 823
Qy 601 GAGTATTTACGTAATTTTGAATATACATGATGATTCCTGATTAACGCTTCTGCGCTG 660
Db 824 GAGTATTTACGTAATTTTGAATATACATGATGATTCCTGATTAACGCTTCTGCGCTG 883
Qy 661 GTGGCTCAAGCACTGGGATGACAGAGAAACAGATTTAAGGAAATGCTGATCCGCG 720
Db 884 GTGGCTCAAGCACTGGGATGACAGAGAAACAGATTTAAGGAAATGCTGATCCGCG 943
Qy 721 CCAATTCGGGAGCAATGAAATTTCTGCGCTGATCACTCCGAGGACCTGGGACATT 780
Db 944 CCAATTCGGGAGCAATGAAATTTCTGCGCTGATCACTCCGAGGACCTGGGACATT 1003
Qy 781 GTTATGGGTTTGGCCGCAACGCTTCTTACTTTGAATATATGAAACGTGTAAA 840
Db 1004 GTTATGGGTTTGGCCGCAACGCTTCTTACTTTGAATATATGAAACGTGTAAA 1063
Qy 841 GAAATCGGTTACCGGACCGGATGATCCGATCATCATGAACTGCCGGCGACCGTGTTC 900
Db 1064 GAAATCGGTTACCGGACCGGATGATCCGATCATCATGAACTGCCGGCGACCGTGTTC 1123
Qy 901 GATCGGACACAGCAATTTGGCAATAACGATGTCCTTTATTTGAAAGCAAGTGAATC 960
Db 1124 GATCGGACACAGCAATTTGGCAATAACGATGTCCTTTATTTGAAAGCAAGTGAATC 1183
Qy 961 TTAATCGGTAACACAGCAACCGATCGTAAAGCCTTATGAAAGGCAAAATTCCTGA 1020
Db 1184 TTAATCGGTAACACAGCAACCGATCGTAAAGCCTTATGAAAGGCAAAATTCCTGA 1243
Qy 1021 GACAAATCGCATGACCTAGATATTAAGTCAACAGATGAATAATTTGTAAGAAA 1080
Db 1244 GACAAATCGCATGACCTAGATATTAAGTCAACAGATGAATAATTTGTAAGAAA 1303
Qy 1081 AGAATGCAACACCGGTGTCTATATGCGGTGCGCAATTTTCATGTCGCCGACAGCTTTA 1140
Db 1304 AGAATGCAACACCGGTGTCTATATGCGGTGCGCAATTTTCATGTCGCCGACAGCTTTA 1363
Qy 1141 ATTGAAAAATCCAGAAATACAGGTAAGAGCTGTAGC 1182
Db 1364 ATTGAAAAATCCAGAAATACAGGTAAGAGCTGTAGC 1405

RESULT 4
US-09-974-300-2694

```

; Sequence 2694, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Beika, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 10085, 500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2694
; LENGTH: 819
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(819)
; OTHER INFORMATION: n = A,T,C or G
US-09-974-300-2694

```

```

Query Match      46.4%; Score 548.4; DB 3; Length 819;
Best Local Similarity 79.9%; Pred. No. 7,4e-139;
Matches 645; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

```

```

QY 24 CTGTCTCTCATCTGTCATCGAATATTGAAAAAGACACATCAGAAAAACATTGA 83
DB 4 CTGTGTATCTGTGTGGGGATCGCATTTTAAAAAGGCCACACGAAAAATATGA 63
QY 84 TGCCTCCCTGTTCCGGTGAATATTAAAGCATCCGCGAAAAATGACTGTGACAGCT 143
DB 64 TGCCTGCTGCTCCAGTGAATCATCAAGGTATACGCGAAAGTCCAGGTGACAAAT 123
QY 144 GACACCGGAAATTAATGAAGCGGTTCAGAGCTGTGAAAAACAAGAAACGA 203
DB 124 AACACAGGATTAATTAATGAAGCGGTTCAGAGCTGTGAAAAACAAGAAACGA 183
QY 204 TGCAAGATGATTTATCTGGACACACCGAGGAAAAAGCGATTAACGGAACCTCAGGG 263
DB 184 CGCAAGATGATTTATTTGGACACACCGAGGAAAAAGCGATTAACGGAACCTCAGGG 243
QY 264 GCGGAATATCGAGAGCAAAAAAGAGTCAAGAGAAAAAGTGAAGAGAGGGCTTAACGC 323
DB 244 GCGGAATATCGAGAGCAAAAAAGAGTCAAGAGAAAAAGTGAAGAGAGGGCTTAACGC 303
QY 324 GATTGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 383
DB 304 GATTGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 363
QY 384 TCTGAGAGCAATATCGGCGTATGATGATGATGATGATGATGATGATGATGATGATGAT 443
DB 364 GCTTATGAGCTATATTCGGCGTATGATGATGATGATGATGATGATGATGATGATGATGAT 423
QY 444 GCGGACGCTTGAATGAAATGCAAGAGCTTTTACCGCTCAATTCCTTATATGSCCATCT 503
DB 424 ACCGACTTGAATGAAATGCAAGAGCTTTTACCGCTCAATTCCTTATATGSCCATCT 483
QY 504 TGTCTTACAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 563
DB 484 GGTATTTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 543
QY 564 AAAAGTCAATCTGCTGATTAATCAAAAAATTAAGATGATGATGATGATGATGATGATGAT 623
DB 544 AAAAGTCAATCTGCTGATTAATCAAAAAATTAAGATGATGATGATGATGATGATGATGAT 603
QY 624 CATGATATTCCTGATTAACGCTTCTTGGCGCTGGGTGAGCTCAAGCACTGGGATTTGA 683
DB 604 CATGATATTCCTGATTAACGCTTCTTGGCGCTGGGTGAGCTCAAGCGTTGGGATTTGA 663

```

```

QY 684 CGAAGAAACAGATTTTAAGGAATGCTGAATGCGCCCGACATCCGGAGCAATGAGAA 743
DB 664 CGAAGAAACAGATTTTAAGGAATGCTGAATGCGCCCGACATCCGGAGCAATGAGAA 723
QY 744 TCTTCCGCTGATCATGCTCGAGCGAGCTGGGCACTTTGTTAAATGAGTTTCCGCAACGA 803
DB 724 TCTTCCGCTGATCATGCTCGAGCGAGCTGGGCACTTTGTTAAATGAGTTTCCGCAACGA 783
QY 804 CGCTTCTTCACTTTGAATATATGGA 830
DB 784 CGAGCTTCACTTTGAATATATGGA 810

```

RESULT 5

```

US-10-724-972A-814
; Sequence 814, Application US/10724972A
; Publication No. US20040147734A1
; GENERAL INFORMATION:
; APPLICANT: Bush, David
; APPLICANT: Doucette-Stamm, Lynn
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: PAT03-16
; CURRENT APPLICATION NUMBER: US/10/724,972A
; CURRENT FILING DATE: 2003-12-01
; PRIOR APPLICATION NUMBER: 09/450,969
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: 09/134,001
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 7544
; SEQ ID NO 814
; LENGTH: 1164
; TYPE: DNA
; ORGANISM: S.epidermidis
US-10-724-972A-814

```

```

Query Match      28.6%; Score 338.6; DB 7; Length 1164;
Best Local Similarity 59.7%; Pred. No. 1.8e-81;
Matches 588; Conservative 0; Mismatches 394; Indels 3; Gaps 1;

```

```

QY 5 GCTGTATCTATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 61
DB 5 GATTGTATCTATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 64
QY 62 GACGATCATGAGAAAAACATTTGATGCTCCCTGCTTGGGATGATTAATTAACGGCATCCGCG 121
DB 65 AGCGTATGCAAAATGCACTTGAAGAAAAATTCATTAAGATTAATTAATTAATTAATTAAT 124
QY 122 GAAATGATCTGTGACAAAGCTGACAAACCGGAATTAATTAAGAACCGGTTTACAAGACTG 181
DB 125 GAAATGATCTGTGACAAAGCTGACAAACCGGAATTAATTAAGAACCGGTTTACAAGACTG 184
QY 182 TTGAAAAACAACGAGAACATATGAAATGATTTACTGGGACACACCGAGAGAAAAAGC 241
DB 185 TTGAAAAACAACGAGAACATATGAAATGATTTACTGGGACACACCGAGAGAAAAAGC 244
QY 242 CGATTTAAACGAAACCTCAGGCGCGGAAATATCGAGAGCAAAAAAGATCATGAGAGAA 301
DB 245 CAGTATATGAAAAACACAGAGCGCGGAAATATCGAGAGCAAAAAAGATCATGAGAGAA 304
QY 302 CAGTATATGAAAAACACAGAGCGCGGAAATATCGAGAGCAAAAAAGATCATGAGAGAA 361
DB 305 TGGTTAAACAAAAAGCAATGCTTGAATTAATGATGATGATGATGATGATGATGATGATGAT 364
QY 362 AAATCATCTTTCAAGAGAACTTCTGCAAGGCAATATCGGCTGATGATGATGATGATGATGAT 421
DB 365 AAATCATCTTTCAAGAGAACTTCTGCAAGGCAATATCGGCTGATGATGATGATGATGATGATGAT 424

```


Db 1065 GAAATCCTTGGAAAGATATTACATGACCCCTAATGACTTATGCTTGGTGAAGTTGCTT 1124
QY 1104 TGGCGTCGGCATATATTCATGTGCGCGAGAGCTTTAATTGAAAAATCCAGCAATACA 1163
Db 1125 GGGATTTGGAAATATTCATGTGTGGTGAAGATTTTTTGTCTTACCTTGTAAGAAAAATA 1184
QY 1164 GGTAAAGCAG 1173
Db 1185 CTTGTGGCAG 1194

RESULT 7
US-10-501-282-6651/c
; Sequence 6651, Application US/10501282
; Publication No. US20050203280A1
; GENERAL INFORMATION:
; APPLICANT: MCMICHAEL, JOHN CALHOUN
; APPLICANT: ZAGURSKY, ROBERT JOHN
; APPLICANT: RUSSELL, DAVID PARISH
; APPLICANT: FLETCHER, LEAH DIANE
; TITLE OF INVENTION: ALLOTIOCCUS OTTIDIS OPEN READING FRAMES (ORFS) ENCODING
; FILE REFERENCE: AM100780 L2
; CURRENT FILING DATE: 2004-07-09
; PRIOR FILING DATE: 2004-07-09
; PRIOR APPLICATION NUMBER: 60/333,777
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 60/426,742
; PRIOR FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: PCT/US02/36123
; PRIOR FILING DATE: 2002-11-25
; NUMBER OF SEQ ID NOS: 6653
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6651
; LENGTH: 1754382
; TYPE: DNA
; ORGANISM: Alloiococcus otitidis
US-10-501-282-6651

Query Match 26.5%; Score 313.2; DB 9; Length 1754382;
Best Local Similarity 55.1%; Pred. No. 7.7e-73;
Matches 634; Conservative 0; Mismatches 513; Indels 3; Gaps 1;

QY 27 TGGCTGCATCTGTCATCGGCAATATTAGAAAAAGAGACATCAAGAAAAATTGATGTC 86
Db 23405 TGGCTCTTATTGGCTAGGCTTATAGAAAGCAGCTTACGAGAAAAATTATTAATA 23346
QY 87 CCTCCCTGTTCCGGTGAATATTAAAGCGCATCCCGGAAATCGACTGTGACAAAGCTGAC 146
Db 23345 AATGACCTAGCGCATTAATTAATTAATGCGACCAAGGCAAGACAGTGAATCTGTTGGT 23286
QY 147 AACCGAATATTATTAAGAAAGCCGGTTACAAAGCTGTGAAAAACAACAGAAACGATGC 206
Db 23285 TACTGGGGGCTCAACAAGCAGCTGCTACCAAGTTATTGGAAAAACAACAGAACTGAGTC 23226
QY 207 AAGATGATTTCTGTGGACACACCGAAGAAAAAGCGAATTAACGAAACCTCAGGGGCC 266
Db 23225 ACGCATTAATTACTGGAGCAGAGCAAGAAAGAGAGATCGAAAGCCCTTCATCTGGAGCC 23166
QY 267 GAATATCGAGAGCAAAAAAGAGTCATGAGAGAAACAGTAGAAAGAGGGGCTAACGCCAT 326
Db 23165 TAAATTTCTGAACAAATATCAGTTATGAAAAAGCAAAAAAGAGGGGCTGATGCCCT 23106
QY 327 TGTCAATGATGATGATGCTGTTTAAACCAAGATTATCAATCATCTTTACAGAAACTTCT 386
Db 23105 TGTATTAAGATGATGATGCTGTTTAAACCAAGATTATCAATCATCTTTATGAATGATTT 23046
QY 387 GAGAGCAATATTCGGGCTATGTGAATTTTAAAGCACTATGAGATGATGAGGGCC 446
Db 23045 TGAATGAATATTAACGACATGTTAAAGCTTTGAAGCAATATGATATCTTGGGTC 22986
QY 447 GAGCGTTGATGAATATGAGAAAGCTTTACCGCTAACATTCCTTATTAATGACCATCTTGT 506

Db 22385 AACAAATGATCAAAATTCCTGGGCTATGCCAAGACCATCCCAAAATGCGCTTGTGT 22926
QY 507 CATTAAGATAGTATATATCCGAGTTCTTTAAACAAAAAGCAAAAGCAACAA 566
Db 22325 CATCTCAAGATGATATTAAGACTTTTATTAAGAGTACCCAGAGAGAAATTCGCC 22866
QY 567 AGTCATGATGCTGATTAATCAAAAATTACAGATGATATTTACGTAAATTTGAATACAT 626
Db 22865 GGTGAAGTTATACAGAACAGCGCTTATGCAAGATTAATGATGCTTGTGCTATCT 22806
QY 627 GGTATTCCTGATATACCGCTTCTGCGCGTGGGTGTGGCTCAAGCATCTCGCATTTGACA 686
Db 22805 TGTCTCCCAACATCTAGCTGTTCGCTGGCTATGTGCAAAACCTCAAGCTTGACCG 22746
QY 687 AGAAACGATTTTAAGGAATGCTGAATGCGCGCGCAGATCCGGAGCAATGAGAAATTC 746
Db 22745 CGATACCTGCATTAAGATATGCTTCAAGCTGACCTGATCCGGCATTTATGCAATGTT 22686
QY 747 TCCGCTGATCAGTCCGAGCGAGCGCTGGGCACTTT--GTTAATGGCTTGGCGCAACGA 803
Db 22685 TGAATAGGGAGGCTATTTGACAGAAAGTTATCTAGCTTATGCTTGTGCTAATGA 22626
QY 804 CGCTTCTTCTACTTGAATATATGAAAACGTGTAAGAAATCGGTTAACCGACCGATGA 863
Db 22625 ACCCATCTTACATGCTGATCTATGAAAGAGCTCAAGAAAGTGGCTTAATAGGTGAAG 22566
QY 864 TCCGATATCATCATGAACTGCGCGGCAAGCCGTGTGATCGACACAGCAATTCGCAAA 923
Db 22565 CCTATTAATTTTGAATCAATGCGAGAAAGACGAGTGAACCGGCTTAATTTATCACCA 22506
QY 924 TGAGATATGCTTATATTTGAAGCAAGTGAATGATCTTATTCGTGAACACAGAAC 983
Db 22505 AGACTTATTTGCTGAATTAAGATGCAAGGCTATGCGCATTTGACAAATACTGGCAT 22446
QY 984 GATGTAAAGCCTATGAGAAAGCAAAATTCCTGACAGCAAACTGCATGACTAGATA 1043
Db 22445 TATTGAGAGGCTTAAATAGGTAGATCAAGGTCAAGGATTAATTTATCTTTGAGAG 22386
QY 1044 TAAGTCAACAGATGAATTAATGAAATGTTTAAAGAAAGATGACAAACCGTCTATATA 1103
Db 22385 GAAATCCTTGAAGATTAATTAATGACCTTATATGCTTGTGTGAAGTTGCTT 22326
QY 1104 TGGCGTCGGCAATATTCAATGTGCGGAGAGCCCTTAATGAAAAATCCAGAAATCAA 1163
Db 22325 GGGTATTTGAATTAATTAATGATGCTGCGTGAAGATTTTTTGTCTTACCTGATGAATA 22266
QY 1164 GGTAAAGCAG 1173
Db 22265 CTTGTGGCAG 22256

RESULT 8
US-09-974-300-3167
; Sequence 3167, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 10085-500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3167
; LENGTH: 309
; TYPE: DNA
; ORGANISM: Bacillus licheniformis

US-09-974-300-3167

Query Match 15.3%; Score 181; DB 3; Length 309;
 Best Local Similarity 74.1%; Pred. No. 1e-38;
 Matches 229; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 873 CATCATGAAGTCCCGGCGAGACCGTGTGATGGACACAGCAATTCGCAATGAGCTAT 932
 DB 1 CCTTATTTACTCCCGGCGCGAGAGGCTGAGCAACACAGCAAGTTGGCGAAGATGCTCT 60
 QY 933 GCGTATATTTGAAGCAAGTGAATCTTAATCGGTGAACACAGAACCGATCGTAA 992
 DB 61 TCTTATATTTGAAGCAAGTGAATCTTGTGATTTGAGAAAACAGAGCCATGCTCA 120
 QY 993 ACCCTATGAAGAGCAAAATTCCTGACAGCAAACTGATGACCTAGAGTATATGTAAC 1052
 DB 121 AGCATATGAAGCAAGCAAAATTCCTGACAGCAAGCTGTTGATTTGAGCAAAATCAAC 180
 QY 1053 AGATGAATTTATGGAATTTGTTAAAGAAAGATGCAACACCGTGTATATGCGCTCG 1112
 DB 181 GGAGAAATCATGTTCTGCTGAAAAACAGCTTGAGGCGCGGTATTTAGCGAGTGG 240
 QY 1113 CAATATTCATGGGCGGAGAGCTTAATTTGAAAAATCCAGATTAAGTAAAGCA 1172
 DB 241 AATATTCACAGGAGCAGCGAGCTCTCATTTGAAAAATGCAAGATTACAAATTAAGCA 300
 QY 1173 GCTCGTAG 1181
 DB 301 GTTCGTTAG 309

RESULT 9
US-10-282-122A-30762

Sequence 30762, Application US/10282122A
 Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Liangu
 APPLICANT: Zamudio, Carlos
 APPLICANT: Malone, Cheryl
 APPLICANT: Haselbeck, Robert
 APPLICANT: Ohlsen, Karl
 APPLICANT: Zyckind, Judith
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John
 APPLICANT: Carr, Grant
 APPLICANT: Yamamoto, Robert
 APPLICANT: Forsyth, R.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 FILE REFERENCE: ELITRA.034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 78614
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 30762
 LENGTH: 1338
 TYPE: DNA
 ORGANISM: Pasteurella multocida
 US-10-282-122A-30762

Query Match 3.8%; Score 44.4; DB 7; Length 1338;
 Best Local Similarity 54.2%; Pred. No. 0.45;
 Matches 90; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 921 AATGACGATTTGCTTATTTATTTGAAGCAAGTGAATCTTAATCGGTGAACACAGCA 980
 DB 369 AGATCCCTTTTCTGCTCATATTTGAAGATGCAACCTTATCTTTATTTGGGGGCAACTGA 428
 QY 981 ACCGATCGTAAAGGCTTATGAGAAAGCAAAATTCCTGACAGCAAACTGCATGACTGA 1040
 DB 429 AATCCTTCTTCTGATTAATTAATGCTTACTTTGCGAGCCAAAGTATATTTCTTAA 488
 QY 1041 GTATATGTCACAGATGAATTTATGGAATTTGTTAAAGAAAGT 1086
 DB 489 ATCGTTGAGCACACAAAGAAATCACCGAGGTGTTAAAGACACTG 534

RESULT 10

US-10-021-323-10884/c

Sequence 10884, Application US/10021323
 Publication No. US20040123340A1

GENERAL INFORMATION:

APPLICANT: Delkman, Jill
 APPLICANT: Feng, Paul C.C.
 APPLICANT: Fincher, Karen L.
 APPLICANT: Ziegler, Todd E.
 TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
 FILE REFERENCE: 38-21(52274)B
 CURRENT APPLICATION NUMBER: US/10/021,323
 PRIOR FILING DATE: 2001-12-12
 PRIOR APPLICATION NUMBER: US 60/255, 619
 PRIOR FILING DATE: 2000-12-14
 NUMBER OF SEQ ID NOS: 17880
 SEQ ID NO 10884
 LENGTH: 600
 TYPE: DNA
 ORGANISM: Gossypium hirsutum

FEATURE:
 OTHER INFORMATION: Clone ID: LIB3828-032-Q6-K6-H10
 US-10-021-323-10884

Query Match 3.6%; Score 42.6; DB 7; Length 600;
 Best Local Similarity 54.0%; Pred. No. 0.91;
 Matches 87; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 160 ATGAGACCGGTTTACAGACTGTTGAAAAACACAGAAACAGATCAGATGATTTAC 219
 DB 566 AAAAGCGCGAAAAACCAAAAAAGGAACTTAAAGAAAGCGCCCAATTTAAATTTAA 507
 QY 220 TGGGACACACCGAGAAAAAGCGATTTAAACGAAACCTCAGGGCGCGAATTCGAGAG 279
 DB 506 AGGGGAAAAAGAGAAAGGGAAGAAAAAGAAAAAGAAAAAGAAAAAGAGGGGGG 447
 QY 280 CAAAAAGATCATGAGAGAAACATGTAAGAAAGGGGCTTA 320
 DB 446 AAAAGGAAGAAAAAGAAAAAGAAAAAGAAAAAGGGGGA 406

RESULT 11

US-09-814-353-17917/c

Sequence 17917, Application US/09814353
 Publication No. US20030165831A1

GENERAL INFORMATION:

APPLICANT: Lee, John

LENGTH: 6000
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-2278

Query Match 3.5%; Score 41; DB 6; Length 6000;
Best Local Similarity 48.1%; Pred. No. 8.5;
Matches 116; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

QY 428 ATATGGATGTCATGGGCGCCGACGCTGATGAATTCAGAGAGCGTTACCGCTACATTC 487
DB 5183 ATTTGGAGCTTGAGGTGAGAGATTATTGAATTAGAGTTGATGAGTTGTAGTGA 5242
QY 488 CTTAATATGCGCATCTTGTCATTACAGATAGTATATACGAGTTCTTTAAACAAAAG 547
DB 5243 GTTATGATGCGGTATTATTATTATTTGAGTATTAAGTGAATTTGTTTAAAAAT 5302
QY 548 CAAAGACGAAACACAAAAGTCATCTGCTGATCTCAAAAATTACAGATGATATT 607
DB 5303 AAAAAATAAAAATATTAATAATTTGTAATTATTAATATATATAGAAAAATAT 5362
QY 608 TACGTAATTTGATATCATGATGTTCCCGATTAACGCTCTCGCGCTGGGTGCTC 667
DB 5363 ATAAATTTATTATATGATTAATGTTTTTTATATATGTTGAGATTTTGGGTGAGATT 5422
QY 668 A 668
DB 5423 A 5423

RESULT 15

US-10-311-455-1788/C
Sequence 1788, Application US/10311455
Publication No. US20030143606A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Detect
FILE REFERENCE: 5013.1014
CURRENT APPLICATION NUMBER: US/10/311,455
CURRENT FILING DATE: 2002-12-16
PRIOR APPLICATION NUMBER: PCT/EP01/07537
PRIOR FILING DATE: 2001-07-02
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 2424
SEQ ID NO 1788
LENGTH: 7312
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1788

Query Match 3.4%; Score 40; DB 6; Length 7312;
Best Local Similarity 53.1%; Pred. No. 18;
Matches 85; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 958 ATCTTATCGGTAAACAAAGAACGATCGTAAAGCTATGAGAAGCAAAATTCCT 1017
DB 4098 AACTAATCAACAAATAAAACCAACCAACGAAAAAATCTATAAAAAACACACC 4039
QY 1018 GCAGACAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1077
DB 4038 GCAAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 3979
QY 1078 AAAAAGATGACAAACGCTGCTATATATGCGCTCGCAATA 1117

DB 3978 TAAGCGCTCTCCACAAACTCAAAAAAAGCATTAACAAA 3939

Search completed: February 27, 2006, 07:58:23
Job time : 1304.32 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:37:00 ; Search time 712.304 Seconds
(without alignments)
3538.457 Million cell updates/sec

Title: US-10-789-164-1

Perfect score: 1182
Sequence: 1 atggagctggtactatcatat.....aggttaagcagctcgtaagc 1182

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 7209121 seqs, 1066183437 residues

Total number of hits satisfying chosen parameters: 14418242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

Published Applications NA.New.*
1: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
2: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq1.*
7: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq1.*
9: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq2.*
11: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq3.*
12: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq4.*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1182	100.0	1182	US-10-530-083-1	Sequence 1, Appl1
2	338.6	28.0	3240	US-10-793-626-3743	Sequence 3743, Ap
3	246.2	20.8	861	US-10-793-626-2321	Sequence 2321, Ap
4	39.6	3.4	4642	US-10-793-626-4090	Sequence 4090, Ap
5	38.6	3.3	1011	US-10-793-626-3209	Sequence 3209, Ap
6	38.2	3.3	103931	US-11-117-187-193	Sequence 193, Ap
7	38.2	3.2	600	US-11-136-527-412	Sequence 412, Ap
8	38.2	3.2	1352	US-11-136-527-26	Sequence 26, Appl
9	38.2	3.2	191343	US-11-112-908-53	Sequence 53, Appl
10	37.6	3.2	519	US-09-925-065A-499536	Sequence 499536,
11	37.6	3.2	519	US-09-925-065A-499538	Sequence 499538,
12	37.6	3.2	519	US-09-925-065A-499539	Sequence 499539,
13	37.4	3.2	519	US-09-925-065A-585886	Sequence 585886,
14	37.2	3.1	519	US-09-925-065A-499537	Sequence 499537,
15	37.2	3.1	552	US-09-925-065A-319482	Sequence 319482,
16	37.2	3.1	191350	US-10-857-780-4	Sequence 4, Appl1
17	36.4	3.1	602	US-09-925-065A-148850	Sequence 148850,
18	36.4	3.1	2796	US-10-793-626-4335	Sequence 4335, Ap
19	36.2	3.1	612	US-09-925-065A-585885	Sequence 585885,
20	36.2	3.1	647	US-09-925-065A-750167	Sequence 750167,

ALIGNMENTS

21	36.2	3.1	647	6	US-09-925-065A-822601	Sequence 822601,
22	36.2	3.1	1002	12	US-11-098-686-9664	Sequence 9664, Ap
23	36.2	3.1	1883	8	US-10-750-185-54329	Sequence 54329, A
24	36.2	3.1	1883	8	US-10-750-623-54329	Sequence 54329, A
25	36.2	3.1	1457619	12	US-11-098-686-8739	Sequence 8739, Ap
26	36	3.0	1547	6	US-09-925-065A-711928	Sequence 711928,
27	35.8	3.0	551	6	US-09-925-065A-831259	Sequence 831259,
28	35.8	3.0	572	6	US-09-925-065A-775813	Sequence 775813,
29	35.8	3.0	572	6	US-09-925-065A-838214	Sequence 838214,
30	35.8	3.0	584	6	US-09-925-065A-298744	Sequence 298744,
31	35.8	3.0	584	6	US-09-925-065A-298745	Sequence 298745,
32	35.8	3.0	612	6	US-09-925-065A-585887	Sequence 585887,
33	35.6	3.0	564	6	US-09-925-065A-756532	Sequence 756532,
34	35.6	3.0	574	6	US-09-925-065A-390144	Sequence 390144,
35	35.6	3.0	627	6	US-09-925-065A-463913	Sequence 463913,
36	35.6	3.0	5388	12	US-11-077-386-14	Sequence 14, Appl
37	35.6	3.0	5980	12	US-11-080-991-95	Sequence 991, Appl
38	35.6	3.0	6147	8	US-10-955-054A-74	Sequence 74, Appl
39	35.6	3.0	6720	12	US-11-077-386-13	Sequence 13, Appl
40	35.6	3.0	6745	12	US-11-077-386-12	Sequence 12, Appl
41	35.6	3.0	6829	8	US-10-955-054A-87	Sequence 87, Appl
42	35.6	3.0	6829	12	US-11-077-386-11	Sequence 11, Appl
43	35.6	3.0	6829	12	US-11-169-041-56	Sequence 56, Appl
44	35.4	3.0	580	6	US-09-925-065A-567282	Sequence 567282,
45	35.4	3.0	586	6	US-09-925-065A-575631	Sequence 575631,

RESULT 1
US-10-530-083-1
Sequence 1, Application US/10530083
GENERAL INFORMATION:
APPLICANT: Sung, Moon-Hee
APPLICANT: Lee, Jong-Seo
APPLICANT: Jung, Chang-Min
APPLICANT: Hong, Seong-Pyo
APPLICANT: Kim, Chul-Joong
APPLICANT: Park, Sue-nie
APPLICANT: Pyo, Hyun-mi
TITLE OF INVENTION: VECTOR FOR ANTI-HPV VACCINE AND TRANSFORMED MICROORGANISM BY THE
FILE REFERENCE: 4240-119
CURRENT APPLICATION NUMBER: US/10/530,083
CURRENT FILING DATE: 2005-04-01
PRIOR FILING DATE: 2002-10-17
PRIOR APPLICATION NUMBER: KR 10-2002-0063378
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn version 3.2
SEQ ID NO 1
LENGTH: 1182
TYPE: DNA
ORGANISM: Bacillus subtilis
US-10-530-083-1

Query Match 100.0%; Score 1182; DB 8; Length 1182;
Best Local Similarity 100.0%; Pred. No. 2,9e+284;
Matches 1182; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATGGGCTGTTACTATTAAGCTGTGCTGTCTATCTGCTATCGGAATATTAGAAAA	60
DB	1	ATGGGCTGTTACTATTAAGCTGTGCTGTCTATCTGCTATCGGAATATTAGAAAA	60
QY	61	CGACGACATCAGAAAAAATTGATGCTCCCTCGGTGGAATTTAAGCGCATCCGC	120
DB	61	CGACGACATCAGAAAAAATTGATGCTCCCTCGGTGGAATTTAAGCGCATCCGC	120
QY	121	GGAATATCGACTGACAAAGCTGACACCGAATATTATAGAACCGGTTACAAAGACT	180
DB	121	GGAATATCGACTGACAAAGCTGACACCGAATATTATAGAACCGGTTACAAAGACT	180

181 GTTGGAAAAACAAGCAAGATGCAAGATGATTCTGGGACACACCGGAGAAAG 240
181 GTTGGAAAAACAAGCAAGATGCAAGATGATTCTGGGACACACCGGAGAAAG 240
241 CCGATTTAAACGGAAACCTCAGGGGCGGAATATCGAGAGCAAAAAAGAGTCATGAGAA 300
241 CCGATTTAAACGGAAACCTCAGGGGCGGAATATCGAGAGCAAAAAAGAGTCATGAGAA 300
301 ACAGTAGAAAGAGGGGCTTAACGGATTTGTCACTGAATGATGCTTTAACCCAGATTAT 360
301 ACAGTAGAAAGAGGGGCTTAACGGATTTGTCACTGAATGATGCTTTAACCCAGATTAT 360
361 CAAATCATCTTTACAGAGAACTTCTGAGAGCCCAATTCGGGCTGATTTGAAATTTTA 420
361 CAAATCATCTTTACAGAGAACTTCTGAGAGCCCAATTCGGGCTGATTTGAAATTTTA 420
421 GAAGACCATATGATGTCATGGGGCGACGCTTGATGAATTTGCAAGAGCGTTTACCGCT 480
421 GAAGACCATATGATGTCATGGGGCGACGCTTGATGAATTTGCAAGAGCGTTTACCGCT 480
481 ACAATTCCTTATATGCGCATTTTGTCTTACAGATAGTAATATACGAGTTCTTTTAA 540
481 ACAATTCCTTATATGCGCATTTTGTCTTACAGATAGTAATATACGAGTTCTTTTAA 540
541 CAAAAGCAAAAAGCAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAA 600
541 CAAAAGCAAAAAGCAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAA 600
601 GAGTATTTACGTAATTTTGAATACATGATTTCCCTGATTAACGCTTCTGCGCGTGGGT 660
601 GAGTATTTACGTAATTTTGAATACATGATTTCCCTGATTAACGCTTCTGCGCGTGGGT 660
661 GTGGCTCAAGCACTCGGCATTTGACGAAGAAACAGCATTTTAAAGGAAATGCTGATCGCG 720
661 GTGGCTCAAGCACTCGGCATTTGACGAAGAAACAGCATTTTAAAGGAAATGCTGATCGCG 720
721 CCAGATCCGGGAGCAATGAAATTTCTCGGTCATGATCGAGCGAGAGCGTGGGCACTTT 780
721 CCAGATCCGGGAGCAATGAAATTTCTCGGTCATGATCGAGCGAGAGCGTGGGCACTTT 780
781 GTTAATGGGTTTGGCCGAAACGACGCTTCTTCTACTTTGAATATATGAAACGTTTAA 840
781 GTTAATGGGTTTGGCCGAAACGACGCTTCTTCTACTTTGAATATATGAAACGTTTAA 840
841 GAAATCGGTTACCGGACCGATGATCCGATCATCATGAACTGCCGCGCAGACCGTGTG 900
841 GAAATCGGTTACCGGACCGATGATCCGATCATCATGAACTGCCGCGCAGACCGTGTG 900
901 GATCGGACACAGCAATTCGCAAAATGACGTAATGCTTATATTTGAAGCAAGTGAATC 960
901 GATCGGACACAGCAATTCGCAAAATGACGTAATGCTTATATTTGAAGCAAGTGAATC 960
961 TTAATCGGTGAAACAAACAGAACCGATCGTAAAGGCTTATGAAAGGCAAAATTCCTCA 1020
961 TTAATCGGTGAAACAAACAGAACCGATCGTAAAGGCTTATGAAAGGCAAAATTCCTCA 1020
1021 GACAAACTGTCATGACCTTGAAGTATATGTCACAGATGAAATTTGTTAAAGAA 1080
1021 GACAAACTGTCATGACCTTGAAGTATATGTCACAGATGAAATTTGTTAAAGAA 1080
1081 AGAATGCAACACCGGTGATATATGCGGCTGATTTCAATGCGCGCAGAGGCTTTTA 1140
1081 AGAATGCAACACCGGTGATATATGCGGCTGATTTCAATGCGCGCAGAGGCTTTTA 1140
1141 ATTGAAAAATTCACGAATATCAAGGTAAAGCAGCTGTAAGC 1182
1141 ATTGAAAAATTCACGAATATCAAGGTAAAGCAGCTGTAAGC 1182

RESULT 2
US-10-793-626-3743
; Sequence 3743, Application US/10793626

Publication No. US20050255478A1
GENERAL INFORMATION:
APPLICANT: KIMBERLY, WILLIAM JOHN
TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
FILE REFERENCE: P03480US
CURRENT APPLICATION NUMBER: US/10/793,626
CURRENT FILING DATE: 2004-03-04
PRIOR APPLICATION NUMBER: 60/164,258
PRIOR FILING DATE: 1999-11-09
NUMBER OF SEQ ID NOS: 4472
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3743
LENGTH: 3240
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-10-793-626-3743

Query Match 28.6%; Score 338.6; DB 8; Length 3240;
Best Local Similarity 59.7%; Pred. No. 3.3e-74;
Matches 588; Conservative 0; Mismatches 394; Indels 3; Gaps 1;
5 GCTGGTACTCATATAGCCTGTGCTG---TCATACCTGTCATCGAATATTAAGAAAAAC 61
1673 GATTGTTACTATATATAGCGGTGTGCTTTCCTTATCTTGGCTAGGAATTAAGAAAGA 1732
62 GACGACATCAGAAAAACATTGATGCTCCTCCCTGTTGGGTTGAATTTAACGCGATCCGG 121
1733 AGCGTCATGCAAAATCGACTGAAAAAATTCATACGATATTAACATTAAGAAATTCG 1792
122 GAAATGCACTGTGACAAAGCTGCAACCGGAATATTAATGAAGCCGTTACAAAGCTG 181
1793 GGAATCTTACGATTAATCTCGGATGCTTACAGTGTCTTACGTAAGGACAGTATAGAGTT 1832
182 TTGAAAAAACAACAGAAACAGATGCAAGAAATGATTTACTGGGACACACCGGAGGAAAA 241
1853 TTGGTAAACAAACAGAAACAGATGCAAGAAATGATTTACTGGGACAGATTAATTCGTAAG 1972
242 CGATTTAAACGAAACCTCAGGGGCGCAATATCGAGAGCAAAAGAAATGATGAGAA 301
1913 CAGTAATCAGAAACCAACAGGCGCCCAACATTTGAGAGAAACAGATTAATTCGTAAG 1972
302 CAGTAAGAAAGGGGCTTAACCGGATTTGTCAGTGAATGCAATGCGGTTAAACCAATATC 361
1973 TGGTTAAACAAAGAAATGCTTATGATGATGATGATGATGATGATGATGATGATGAT 2032
362 AAATCATCTTTCAGGAAGAACTTCTGAGGCGCAATATCGGCGTATTTGATGATTTTAA 421
2033 AAATTTACTTTCAAAATGATTTTAAAGCAATATCGGTTAAATTTGTTAAATGATGATG 2092
422 AAGACCATATGATGTCATGCGGCGCAACGCTTGATGAATTTGCAAGACGTTTACCGCTA 481
2093 AAGACCATATGATGTCATGCGGCGCAACGCTTGATGAATTTGCAAGACGTTTACCGCTA 481
482 CAATTCCTTAATATGCGCATTTTGTCAATACAGATAGTAATATACGAGTTCTTTTAAAC 541
2153 CAATTCCTTAATATGCGCATTTTGTCAATACAGATAGTAATATACGAGTTCTTTTAAAC 541
542 AAAAAAGCAAAAGCAAAACAAAGTATCATTTGCTGATTAACCAAAATTCACAGATG 601
2213 AGAAGCTTAAAGCGTAAATTCAGAACTCATTTGTTTGAATTAAGAGCTATACAGAAAT 2272
602 AGTATTTACGTAATTTGTAATACATGATTTTCCCTGTAACGCTTCTGCGCGTGGGTG 661
2273 CATATTTACGGAAGTGCATATTTAGATTTCTGTAATATGATGATGATGATGATGATG 2332
662 TGGCTCAAGCACTGGGCAATTAAGCAAGAAACAGATTTAAAGGAAATGCTGAATCGCCG 721
2333 TAGCGCAAGCAGTGTGTTAGATGAAAGAACTGATTAACAGGATGATGATTAATCAACAG 2392
722 CAGATCCGGGAGCAATGAAATTTCTCCGCTGATCAATGTCGAGGAGCGTGGCACTTTG 781

Accession	Sequence	Position
Db	CCGATCCACGGTGCCTGTAGATTTAAATATATTTCCATGCAAAATCGCAAAAAAATGATTTTG	2452
Qy	TTATTCGGTTTCCGCAACAGACCTCTCTTACTTTGATATATATGAAACGCTATAAG	841
Db	TTAATGCACTTGCCTGTATATGACCGCAGCTTACAAAACGATTTTAAATAAATGAGAT	2512
Qy	AAATCGGTTACCCGACCGATATCCGATCATCATGATGAACCTGCCGCGACACCGCTGCG	901
Db	CATATTAATTAATCCATACGATTAAGAAAAAATATCATTTCTAATTTGCTTCAAGTAAAGGCTG	2572
Qy	ATCGGACACAGCAATTCGCAATGACGTATTCCTTATATTTGAACCAAGTGAATGATCT	961
Db	ATAGAACCAACCTCTTTGTGTATTACTTTTATGGAAGTGCATTAACGATGTTCTCATTT	2632
Qy	TAATCGGTGAACCAACAGACCGAT	986
Db	GTACAGAAAAAGTACACAATGCT	2657

RESULT 3
US-10-793-626-2321

```

Sequence 2321, Application US/10793626
Publication No. US20050255478A1
GENERAL INFORMATION:
APPLICANT: KIMMERLY, WILLIAM JOHN
TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
FILE REFERENCE: PU3480US
CURRENT APPLICATION NUMBER: US/10/793,626
CURRENT FILING DATE: 2004-03-04
PRIOR APPLICATION NUMBER: 60/164,258
PRIOR FILING DATE: 1999-11-09
NUMBER OF SEQ ID NOS: 4472
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2321
LENGTH: 861
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-10-793-626-2321

```

Query Match	20.8%	Score 246.2	DB 8	Length 861
Best Local Similarity	60.0%	Pred. No. 2e-51		
Matches 410; Conservative	0;	Mismatches 273;	Indels 0;	Gaps 0

Oy 304 GTAGAAAGGGGCTTACCGCATTTGTCACTGATGATGCATGGCTGTTAAACCCAGATTATCA 363
 Db. 4 GTTAAACAAAGGCAATGCCCTAGTTATGATGTATGGCTGTAAATCCAGATTATCA 63
 Oy 364 ATCATCTTTCAAGAAACCTTCTGCAGGCCAATATGCGCGTCAATGTGAATGTTTAA 423
 Db 64 ATTACCTTTCAAAATGATTAGTAAAGCAAAATATGCGTGTATTTGTAATGTAGTGAA 123
 Oy 424 GACCATATGATGTCAATGGGCGGACGCTTGATGAAATTGCAGAACGTTTACCGTACA 483
 Db 124 GACCATATGATGTCTTAAAGCCAGACATTTAAAGATGTAGCGCACCTTTTACGTCAACA 183
 Oy 484 ATTCTTATATGGCCATCTGTGCACTTACAGATGATGAATATACGAGTTCTTTAACA 543
 Db 184 ATTCATATPACGGGAATTAGTTGTAATGAAGAATTAATACTATATCTTGTTCGAAAG 243
 Oy 544 AAAGCAAAAGACGAACAACAAGATCATCTTGTGATTAACCAAAAATTACAGATGAG 603
 Db 244 GAGCTAAAAAGGCTAATTCAGAACTCATTTGTGTATTAABAACGTCAATACAGATCA 303
 Oy 604 TATTACGTAAATTTGATACATAGTATTCCTGATPACGCTTCTGTGGCGTGGGATG 663
 Db 304 TATTTCGGAAGTTCATTTATAGTATTTCTGATATATGTACTATATGTGTAGGAATA 363
 Oy 664 GCTCAAGCACTGGCATTTGACGAAGAACGCAATTTAAGGGAATGCTGAATGCGCCCA 723

Db	364	GCGCAGAGCATGGTGTGATGAAGAAACCTGATATCAAGATATGTTAAATGACACGAGCC	423
Qy	724	GATCCGGAGACATAGAAATTCCTCCCGTCGATCACTCCGAGGAGCGCTG8GGACCTTGT	783
Db	424	GATCCAGGTGCTGTAGATTAAATATATTTTCATGCAAAATCGACAAAAAATATGATTTGTT	483
Qy	784	AATGGATTGCGCGAAGCAGCGCTTCTTCTACTTGAATATATGGAACGTTAAGAA	843
Db	484	AATGATTTGCGTCTTAATGAACCGCAGTCTACAAAAGCGATTTTAAATAAAGTGATCA	543
Qy	844	ATCGGTTACCGACCGATGATCCGATCATCATGAACTGCGCGCAGACCGTGTGCAT	903
Db	544	TATATTTATCCATACGATAGAAAAATATCATTTCTCAATTTGCGTTCAAGTAGGGTGTAT	603
Qy	904	CGGACACAGCAATTCGCAATGACGATTTGCTTATTTATGGAAGCAAGTGAACCTGATCTTA	963
Db	604	AGAACACAACTCTTTGTGATTAACCTTTTAAAGTGAAGTGATTAAGATGCTTCATATTGT	663
Qy	964	ATCGGTGAAGAACAGAACCGAT	986
Db	664	ACAGGAAAAGTACACCAATGTG	686

RESULT 4
US-10-793-626-4090

```

: Sequence 4090, Application US/10793626
: Publication No. US20050255478A1
:
: GENERAL INFORMATION:
: APPLICANT: KIMMERLY, WILLIAM JOHN
: TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND
: FILE REFERENCE: PU3480US
: CURRENT APPLICATION NUMBER: US/10/793,626
: CURRENT FILING DATE: 2004-03-04
: PRIOR APPLICATION NUMBER: 60/164,258
: PRIOR FILING DATE: 1999-11-09
: NUMBER OF SEQ ID NOS: 4472
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 4090
:
: LENGTH: 4642
:
: TYPE: DNA
:
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: synthetic
:
: OTHER INFORMATION: nucleic acid sequence
:
: US-10-793-626-4090

```

Query Match	3.4%	Score 39.6	DB 8	Length 4642
Best Local Similarity	43.6%	Pred. No. 9.6		
Matches 177	Conservative	0	Mismatches 229	Indels 0
			Gaps	0

Qy	237	AAAGCCGATTAAACGGAAACCTCAGGGGCCGATATTCGAGAGCAAAAAGATCATCAG	296
Db	2052	AATGGCAATTAAATGAGCAATTAAATGGTTTGGTGAATTTGGTCGTTTACGATTCAGAA	2111
Qy	297	AGAAACAGTGAAGAGGGGCTTACCGCATTTGTCAATGATGCAATGGCTGTTAACCCGAA	356
Db	2112	AATTCAGATGTGAGAGGCTTTGAAGTGTGCAGTTAACGATTAACGATGACGATAT	2171
Qy	357	TTATCAAAATCATCTTTCAGAAAGACTTCGAGGCCAAATATCGGCTCATTTGTAATGT	416
Db	2172	GTTAGCTATTTATTTAAATACGATACATATGCAAGTGGTTTCACTGGAGAAGTTGAAGT	2233
Qy	417	TTTAGAAGACCATATGATGTCATGAGGGCCGACGCTTAGATGAATTCGAAAGCCTTAC	476
Db	2232	TATCGAAGGTGATTCCTGGTGAAACGATAAAGAAATTAATCAATTCGATGAAACCGAATGC	2292
Qy	477	CGCTACAAATTCCTTTAATATGCCATCTTGTCAATTACAGATGTGAATTAACGAGTCTCT	536
Db	2292	TGGTAAATTATACATGGGGCATTTTAGATATCGACGTAGATTAGATGATCTGTTTCTA	2355
Qy	537	TAAACAAAAGCAAAAGAACGAAACACAAAGTCATCTATTCGTGCTAACTCAAAATTTAC	596
Db	2352	TACTGTATTAAGAAAAAGCAAGAGTCACATCGATCGAGTGGTTAAAAAGATTAAATCTC	2411

OY 598 GATGAGTATTACGTAATTTGATATACATGATTCCTGATACGCT 645
 DB 239 GGTAAGGATTGCTTARATTTTATCTTGAGGCCAAGATTAGTCT 286

RESULT 15

US-09-925-065A-319482/c
 ; Sequence 319482, Application US/09925065A
 ; Publication No. US20040181048A1
 ; GENERAL INFORMATION:

APPLICANT: Wang, David G.
 TITLE OF INVENTION: Identification and Mapping of Single
 ; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
 ; FILE REFERENCE: 108827.135
 ; CURRENT APPLICATION NUMBER: US/09/925,065A
 ; CURRENT FILING DATE: 2001-08-08
 ; PRIOR APPLICATION NUMBER: US 60/243,096
 ; PRIOR FILING DATE: 2000-10-24
 ; PRIOR APPLICATION NUMBER: US 60/252,147
 ; PRIOR FILING DATE: 2000-11-20
 ; PRIOR APPLICATION NUMBER: US 60/250,092
 ; PRIOR FILING DATE: 2000-11-30
 ; PRIOR APPLICATION NUMBER: US 60/261,766
 ; PRIOR FILING DATE: 2001-01-16
 ; PRIOR APPLICATION NUMBER: US 60/289,846
 ; PRIOR FILING DATE: 2001-05-09
 ; NUMBER OF SEQ ID NOS: 957086
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 319482
 ; LENGTH: 552
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-925-065A-319482

Query Match 3.1%; Score 37.2; DB 6; Length 552;

Best Local Similarity 48.6%; Pred. No. 17;
 Matches 102; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

OY 408 TGTGAATGTTTGAAGACCATATGATGATGCGGCGACGCTTGATGAATTCAGA 467
 DB 452 TATGACCTAATATACATTTTCAGATATACAAAGCAACATTTACATGATGATGACG 393
 OY 468 AGCGTTACCGCTACATTCCTTATAATGCGCATCTTGCTATTACAGATATGATATAC 527
 DB 392 AGATAAATTTAAACTTTAGCTGAATGTTTATGTTTATACAGACACTGAAAAAT 333
 OY 528 CGAGTCTTTAAACAAAAGCAAAAGACGAAACAAAAGTCATCTGCTGATAC 587
 DB 332 GAAGTAGACAAAATGTAAACATTATATGACAAATTTGAAACATTAACAAAGCTTACACTG 273
 OY 588 AAAAATTACAGATGATATTACGTAATTT 617
 DB 272 ATTAAATAAAGTTATTTTAAATGT 243

Search completed: February 27, 2006, 11:38:27
 Job time : 713.304 secs

THIS PAGE BLANK (USPTO)


```
Db 305 TGGTTAAACAAAAGCAATATGCCCTAGTTAATGAGTATGCGCTTAATCAAGATTATC 364
Qy 362 AAATCATCTTTTCAGAGAACTTTCGACGGCCATATGCGGCTCATTTGTGAATGTTTAA 421
Db 365 AAATTCACCTTTCAAAATATTAGTAAAGCAAAATCGGTGTAATTTGTATGTGAAGG 424
Qy 422 AAGACCATATGATGTCATGCGGCGCGCATTTGATGAATTTGCAAGACGTTTACCGCTA 481
Db 425 AAGACCATATGATGTCATGCGGCGCGCATTTGATGAATTTGCAAGACGTTTACCGCA 484
Qy 482 CAATTCCTTAATATGCGCATCTTGTCAATTAAGATGTAATATACCGAGTTCTTTAAAC 541
Db 485 CAATTCCTTAATATGCGCAAAATTTAGTTTAAGAAATACATATCTAGTTTCTTTGAA 544
Qy 542 AAAAAAGCAAAAGCAAAACACAAAGTCATCTGCTGATTACTCAAAAAATTACAGATG 601
Db 545 AAGAAAGCTAAAGAGCTAATTCAGAACTCATGTTGTGATTAAGACGTCATACCAAGAT 604
Qy 602 AGTATTTAGTAATTTTGAATATCATGATATTCCTGATTAAGCCTTCTCGGCGCTGGGTG 661
Db 605 CATATTTAGGGAAGTTTCATTTTATTTAGTATTTCTGATTAATGATGATTTGTTAGAA 664
Qy 662 TGGCTCAAGCACTCGGCAATTCAGAAAGCAAGCAATTTAAGGGAATGCTGAATGCGCCG 721
Db 665 TAGCGCAAGCAAGTTGCTGATGAATGAAGAAACGTCATTAAGGATGTTAATGACACAG 724
Qy 722 CAGATCCGGAGACATGAATTTCTCCGCTGATGATGTCGACGACGCTGGGCACTTTG 781
Db 725 CCGATCCAGGTGCTGTTGAATTAATTAATTTCCATGCAAAATCGCAAAAAATGATTTG 784
Qy 782 TTAATGGGTTTGGCGCAACGCGCTTCTTCACTTTGAATTAATGAAGCGTGAAG 841
Db 785 TTAATGCAATTCCTGCTCTATGAAACCGCAGCTTACAAAAGCAATTTAAATTAAGTGAA 844
Qy 842 AAATCGTTAACCAGCCGATGATCCGATCATCATCATGAACTCCGCGCAGACCGTGTG 901
Db 845 CATATTAATTCATACGATTAAGAAATATATCATTTCTCAATGTGCTTCAAGTTAGGTTG 904
Qy 902 ATCGACACAGCAATTCGCAATGACGATTCCTTATATTTGAAGCAAGTGAACGATCT 961
Db 905 ATAGAACAACAACCTCTTTGTTGTAATCTTTTATGGTGAAGTCATTAAGATGTTCTCATTT 964
Qy 962 TAATCGTGAACAACAGAACCGAT 986
Db 965 GTACAGAAAAAGTACACAAATGCT 989
```

RESULT 2

```
US-09-710-279-3743
; Sequence 3743; Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMBERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3743
; LENGTH: 3240
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-09-710-279-3743
```

Query Match 28.6%; Score 338.6; DB 3; Length 3240;
Best Local Similarity 59.7%; Pred. No. 1.7e-86;

```
Matches 588; Conservative 0; Mismatches 394; Indels 3; Gaps 1;
Qy 5 GCTGTTACTCATATATGCGTGTGCTG---TCATACTGTCATCGAATATTTAGAAAC 61
Db 1673 GATGTTACTCATATATGCGTGTGCTGCTTATTTCTTGGCTAGAAATTAAGAGAA 1732
Qy 62 GACGACATCAAGAAAATGATGCTCCCTGCTTCCGGTGAATATTAACGGCATCCGC 121
Db 1733 AGCGTCATGCAAAATGCACTGAAAATAATTCATAGATATTAACATTAATGAATTCGTG 1792
Qy 122 GAAATGACTGTGACAAAGCTGCAACCGGAATATTAATGAAGCCGTTTACAAGCTG 181
Db 1793 GGAATCTACGATTAATCCGATGCTTACAGTGTCTACGCTGAGACACAGATTAAGCTTG 1852
Qy 182 TTGAAAAAACAAGCAAGATGCAAGATGATTTACTGAGGACACACCGAGGAAAG 241
Db 1853 TTGTTAAGACACAGAAACGATGCAAGATGCTCTTATTTGGTTTACTGAAAAGGAATATC 1912
Qy 242 CGATTTAAACGAAACCTCAGGGGCGCAATATCGAGAGCAAAAAGAAAGTCAAGAGAA 301
Db 1913 CAGTAATCAGAAAACCAAGGGCGCAACATTTGAGAAACAAGATATTTATTCGTAAGG 1972
Qy 302 CAGTAAGAAAGAGGGCTTAAGCGCATTTGTCACTGATGATGCTGTTAACCAGATATC 361
Db 1973 TGGTTAAACAAAAGCAAAATGCCCTAGTATAGATGATGCTGTAATCCAGATATC 2032
Qy 362 AAATCATCTTTCAAGAAACCTTCTGACAGCCATATCGGGCTCATTTGTGAATGTTTAA 421
Db 2033 AAATTCCTTTCAAAATGATTTAGTAAAGCAATATCGGTATATTTGTTAATGATGG 2092
Qy 422 AAGACCATATGATGTCATGCGGCGCGACGCTTGAATGAAATTCAGAAAGCGTTTACGCTA 481
Db 2093 AAGACCATATGATGTCATGAGACGCAACCTTAAGATGTAAGCCAAAGCTTTACTGCA 2152
Qy 482 CAATTCCTTAATATGCGCATTTTGTCAATACAGTATGTAATATACCGAGTTCTTTAAAC 541
Db 2153 CAATTCATATTAAGGGAATTTAGTTGTAATGAAGAAATCTAATACATTAAGTTTCTTCA 2212
Qy 542 AAAAAAGCAAAAGCAAAACCAAAAGTCATCTTGGCTGATTACTCAAAAAATTACAGATG 601
Db 2213 AAGAAAGCTAAAGAGCTAATTCAGAACTCATTTGTTGATTAAGATTAAGACGTCATACCAAGAT 2272
Qy 602 AGTATTTACGTAAATTTTGAATATCATGATATTCCTGTAACGCTTCTTGGCGCTGGGTG 661
Db 2273 CATATTTACGGAAGTTGATTAATTTAGTATTTCTGTAATATGATGATGTTAGTAA 2332
Qy 662 TGGCTCAAGCACTCGGCAATTCAGAAAGCAAGCAATTTAAGGGAATGCTGAATGCGCCG 721
Db 2333 TAGCGCAAGCAAGTTGGTGTGATGAAGAACTGCAATTAACAAGATGTTAATACACAG 2392
Qy 722 CAGATCCGGAGCAATGAGAAATTTCTCCGCTGATCAATCCGAGGAGCGCTGGCACTTTG 781
Db 2393 CCGATCCAGGTGCTGTTGAATTAATTAATTTCCATGCAAAATCGCAAAAAATGTAATTTG 2452
Qy 782 TTAATGGGTTTGGCGCAACGACGCTTCTTCACTTTGAATATATGAAAACGTTAAAG 841
Db 2453 TTAATGCAATTCGCTGCTAATTAAGAACGAGTACAAAAGGATTTTAAATTAATGGAAT 2512
Qy 842 AAATCGTTAACCAGCCGATGATCCGATCATCATCATGAAGTCCGCGCAGACCGTGTG 901
Db 2513 CATATTAATTCATACGATTAAGAAATTAATCAATTTCTCAATTTGCTTCAAGATGGGTG 2572
Qy 902 ATCGACACAGCAATTCGCAATGACGATTTGCTTATATTTGAAGCAAGTGAATGATCT 961
Db 2573 ATAGAACAACAACCTCTTTGTTGATTAATCTTTTATGGTGAAGTCGATTAAGATGTTCTCATTT 2632
Qy 962 TAATCGTGAACAACAGAACCGAT 986
Db 2633 GTACAGAAAAAGTACACAAATGCT 2657
```

RESULT 3

US-09-710-279-2321

```

: Sequence 2321, Application US/09110279
: Patent No. 6703492
:
: GENERAL INFORMATION:
: APPLICANT: KIMMERLY, WILLIAM JOHN
: TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
: FILE REFERENCE: PUS480US
: CURRENT APPLICATION NUMBER: US/09/710,279
: CURRENT FILING DATE: 2000-11-09
: PRIOR APPLICATION NUMBER: 60/164,258
: PRIOR FILING DATE: 1999-11-09
: NUMBER OF SEQ ID NOS: 4472
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 2321
:
: LENGTH: 861.
:
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: synthetic
: OTHER INFORMATION: nucleic acid sequence
: US-09-710-279-2321

```

RESULT 4
 US-08-232-463-14/c
 : Sequence 14, Application US/08232463
 Patent No. 5670367
 : GENERAL INFORMATION:
 APPLICANT: DORNER, F.
 APPLICANT: SCHIEFLINGER, F.
 APPLICANT: FALKNER, F. G.
 TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
 NUMBER OF SEQUENCES: 52
 : CORRESPONDENCE ADDRESS:
 ADDRESSEE: Foley & Lardner
 STREET: 1800 Diagonal Road, Suite 500
 CITY: Alexandria
 STATE: VA
 COUNTRY: USA
 ZIP: 22313-0299
 : COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/232.463
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/07/935.313
 FILING DATE:
 APPLICATION NUMBER: EP 91 114 300.6
 FILING DATE: 26-AUG-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: BENT, Stephen A.
 REGISTRATION NUMBER: 29,768
 REFERENCE/DOCKET NUMBER: 30472/114 IMMU
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703)836-9300
 TELEFAX: (703)683-4109
 TELEX: 899149
 INFORMATION FOR SEQ ID NO: 14:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 7218 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 CLONE: PTZgpt-F18
 US-08-232-463-14

Oy 371 TTCAGAGAACTTCTGCGAGCAATATCGGCTCATTTGTAAGTTTAGAGCCATA 430
Db 1155 RRR 1096
Oy 431 TGGATGTCATGGGCGGCGCTTGATGAATGCA 465
Db 1095 RRR 1061

RESULT 5
US-09-248-796A-2670
Sequence 2670, Application US/09248796A
Patent No. 6747137
GENERAL INFORMATION:

APPLICANT: Keith Weinstock et al
TITLE OF INVENTION: NOCLETIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
FILE REFERENCE: 107196.132
CURRENT FILING DATE: US/09/248,796A
PRIOR APPLICATION NUMBER: US 60/074,725
PRIOR FILING DATE: 1998-02-13
PRIOR APPLICATION NUMBER: US 60/096,409
PRIOR FILING DATE: 1998-08-13
NUMBER OF SEQ ID NOS: 28208
SEQ ID NO 2670
LENGTH: 2352
TYPE: DNA
ORGANISM: Candida albicans
US-09-248-796A-2670

Query Match 3.8%; Score 45.4; DB 3; Length 2352;
Best Local Similarity 51.8%; Pred. No. 0.015; Mismatches 96; Indels 0; Gaps 0;

Matches 103; Conservative 0; Mismatches 96; Indels 0; Gaps 0;
Oy 500 ATCTGTGATTCAGATAGTGAATATACCGAGTCTTTAAACAAAGCAAAAGCAAA 559
Db 587 ATATTGTCTATGCAATATATATGTAAGCAATATTAGCTTAACACAGCTAAACAGAA 646
Oy 560 ACACAAAGTCACTCTGCTGATACTCAAAATTCAGATGATTAATTAATTTTG 619
Db 647 ATCAAAATCAATTACTGAGCTAAACCAATGTTCACTGAAGTGGTCTTCCATG 706
Oy 620 AATACATGATTCCTCTGATACGCTTCTCTGCGCTGGCTGCTCAAGACTCGGCA 679
Db 707 AATCTATTCAGATATCTAAACCAATTAATCACTAATTAATTAATGCTGCTGACCTCCAG 766
Oy 680 TTGACGAGAAACAGCATT 698
Db 767 GTACTGTAAACATCATTT 785

RESULT 6
US-09-806-708B-22/c
Sequence 22, Application US/09806708B
Patent No. 6784342
GENERAL INFORMATION:
APPLICANT: The University of British Columbia
TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
FILE REFERENCE: 4810-58741
CURRENT FILING DATE: US/09/806,708B
PRIOR APPLICATION NUMBER: US 60/147,133
PRIOR FILING DATE: 1999-08-04
NUMBER OF SEQ ID NOS: 23
SOFTWARE: PatentIn version 3.0
SEQ ID NO 22
LENGTH: 1141
TYPE: DNA
ORGANISM: Artificial sequence
FEATURE:
NAME/KEY: promoter

LOCATION: (1)..(1141)
OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. PAB1 promoters
US-09-806-708B-22

Query Match 3.4%; Score 40.6; DB 3; Length 1141;
Best Local Similarity 9.6%; Pred. No. 0.26;
Matches 54; Conservative 241; Mismatches 266; Indels 2; Gaps 1;

Oy 13 CTCATTAAGCCTGCTGCTCATCTGATCGATCGAATATTAAGAAAACGACGATCAG 72
Db 619 MTYAKKVKMRDPTTCTTVDVADSVWYXWMMCRDPTTIRNTYCKSVXHSYVWENNA 560
Oy 73 AAAAACTGATGCGCTCCCTGTTGGGTAATTTAAACGCGATCCGGGAAAATGCACT 132
Db 559 MWYRRYARWWSWABRTTNNMMWSGBVRRMAGTVMWMMNNNNNDTRYYMMKMAR 500
Oy 133 GTGCAAGGCTGACACCGGATTTATATGAAAGCCGTTTCAAGACTGTGGAAAAACA 192
Db 499 BTTVYDSMCVAKSMWGNWRAKMMWAAANDAGAMDHTYWGNNNTMMRRRAKMM 440
Oy 193 ACAGAAACAGATGCAAGATGATTACTGGGACACACCGAGAAAACCGATTAAACGG 252
Db 439 AWCRRATCCNNNNNRACVNHKMMRWTKMKKACNNNNNRKAMRYRVAMWTSRDTN 380
Oy 253 AAACCTGAGGCGCGAATATCGAGACAAAGAACTCATGAGAAACAGTAGAAGA 312
Db 379 TDMMWMTSDMBMHVTVDTYMMRAMNNNNNNNNNNBCKTSMWMMMDHNTHTCTGNNTWG 320
Oy 313 GGGGCTAACGGATTCAGTGAATGCAATGCGCTTTAAACCAATTTCAATATCTTT 372
Db 319 SAYMAAMSMWAAASBIVYUWCMWNAKAKVATKMMWYTTDRYVSANNTGVAMWNRW 260
Oy 373 CAGGAAGACTTCGACGCGCAATATCGAGCAATGATGATGTTTGAAGAAGCATATG 432
Db 259 DTAWTBRNKTYAYAVWYI--BMYGKHMBWRABHRSMMWVKCNKMYVSWHY 202
Oy 433 GATGTCATGGGCGCGAATATCGAATGTAATTCAGAACGTTTACCGCTACATTCCTAT 492
Db 201 HAMRYBKABAVGNNMMKMBMAHHHCATNNNNMMWVYMHMMHKKXAAWTNNKTAB 142
Oy 493 AATGGCATCTTTCATTAACATGATATATCCGAGTCTTTAAACAAAGCAAAA 552
Db 141 RDBHBAHVKTWYRVYUWCMWNAKAKVATKMMWYTTDRYVSANNTGVAMWNRW 82
Oy 553 GAAGAAACAAAGCATCAT 575
Db 81 CMWYSMNNRWYRMGRKTYAW 59

RESULT 7
US-09-806-708B-22
Sequence 22, Application US/09806708B
Patent No. 6784342
GENERAL INFORMATION:
APPLICANT: The University of British Columbia
TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
FILE REFERENCE: 4810-58741
CURRENT FILING DATE: US/09/806,708B
PRIOR APPLICATION NUMBER: US 60/147,133
PRIOR FILING DATE: 1999-08-04
NUMBER OF SEQ ID NOS: 23
SOFTWARE: PatentIn version 3.0
SEQ ID NO 22
LENGTH: 1141
TYPE: DNA
ORGANISM: Artificial sequence
FEATURE:
NAME/KEY: promoter
LOCATION: (1)..(1141)
OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. PAB1 promoters
US-09-806-708B-22

Query Match 3.4%; Score 40.4; DB 3; Length 1141;
Best Local Similarity 9.7%; Pred. No. 0.3;
Matches 71; Conservative 285; Mismatches 373; Indels 2; Gaps 1;

```
406 ATTGGAATGTTTGAAGACCATATGATGATGGGCGACGCTTGATGAATTCGA 465
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
71 WYNNKSRWKKWYKQKWBKANTSBTHARKKMKTRAYBMTNNKKGKWRHRYRW 130
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
466 GAAGCGTTTACCGCTACATTCCTTATAATGACCATCTTGCTATTAAGATGAT 525
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
131 RMBDTVDHYYTANNNAMTTTCMDKDKTRMMWKNNATGDDDTKYMNNNGCB 190
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
526 ACCGAGTCTTTAAACAAAGCAAAAGACAAACAAAGTATCATCTGCTATAC 585
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
191 TTTWVVRYXTDMSBKNNYGBMWKWSYDVYTTWVWMDMCKRYARWRTGRMKN 250
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
566 TCAAAAATTACAGATGATATTAGTAAATTTGAATACATGATATCCCTGATA 645
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
251 YVABMTAARRRYNNWBTBAAAYRRKTMNNNNNNKAKMKRYGMNRABVNSTCT 310
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
646 TCTCTGGCGCTGGTGTGCTCAAGCACTCGGCATTGACGAAAGAAACAGCAT 705
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
311 SKTTVTRTSQWANNCRAGDANKDHKWKYSAAAGVYNNNNNNNNNTYKKAHBA 370
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
706 ATGCTGATGCGCCCGCCAGATCCGGAGCAATGAGAAATTCCTCCGCTGAT 765
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
371 HSAWKMHANAAHYSKWTBYRKTMVNNNGTTWKKMMWAMYMDMDMBGTYN 430
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
766 GAGCCG--GGCACTTTGTAATGGTTGGCGCAACGACGCTTCTTACTTGA 823
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
431 GRTTYGWTNKKKKMYKWKANNCKRAMDHKTCTHNNNTWKMTYNNNCYKSM 490
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
824 TATGGAACGTTAAAGAAATCGTTACCGACCGATGATCGATCATCATGAACT 883
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
491 KHRBAAAYTYMMWRRYAAHANNNNWYMWKACITWYKVBOSKMMNYAAVY 550
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
884 GCCCGCCAGACCGTGCATCGGACACAGCATTCGCAATGACGATTCCTTAT 943
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
551 YTSRYRMKTNNNSWRSDTRSGRANNYARABHYGYKNTRWBSHTBHBRA 610
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
944 AAGCAAGTACATGATCTTAATCGGTGAACAACAAGCAAGATCGTAAAG 1003
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
611 WMBMVBACHCKWAKYKAKKYTAGAGSNNNNNNNNNNNNNNNNNNATCA 670
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1004 AAGGCAAAATTCCTGACGAACTGACCTAGAGTATAAGTCAAGATGAAT 1063
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
671 MANAKYKYYKBAANAYTHANNWGCNNATDTRTMMNNNNNNNAGTMM 730
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1064 TCGAATGTTAAAGAAAGATGACACCGTGTCAATATGCGCGCATATTC 1123
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
731 ASAAKRYAAAYKAKKHWKAMKMGWHDAAABTDKRNNGAYTKYTTNN 790
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1124 GTGCCGACAG 1134
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 791 GVVNTTAARDG 801
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 8

US-09-134-001C-2676
Sequence 2676, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674

SEQ ID NO 2676
LENGTH: 1047
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2676

Query Match 3.4%; Score 39.6; DB 3; Length 1047;
Best Local Similarity 43.6%; Pred. No. 0.48;
Matches 177; Conservative 0; Mismatches 229; Indels 0; Gaps 0;

```
237 AAAGCGATTAAACGGAACCTCAGGGCCGAATATCGAGACGAAAGATCATAG 296
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
36 AATGCAATTAAAGATGAATTAATGTTTGTAAGATTTGCTTTAGCATTCAGAG 95
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
237 AGAAACATGAAAGAGGGCGTTACCGCATTTGTACGTAATGCAATGGCTTTAC 356
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
96 AATTCAGATGAGAGGCTTTGAAAGTAGTTCAGTTAACAATTACAGATGACAT 155
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
357 TTATCAATCATCTTTCAGAAAGACTTTCAGAGCCGAATATCGGCGTCAATTG 416
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
156 GTTAGCTATTATTATAATAGACTATGCAAGTGTGTTCACTGAGAAAGTTGAG 215
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
417 TTTAGAACATATGATGATGATGAGGCGCCAGCGCTGATGAAATTCAGAA 476
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
216 TATCGAAGTGGATTCCGTGTAACGTTAAAGAAATTAATCATTCGATGAA 275
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
477 CGCTCAATTCCTTATATGGCAATCTTGCTATTACAGATGTAATACGAGTCT 536
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
276 TGGTAAATTAACATGGGCGATTAGATATGACGATGATAGATGATGCTGTT 335
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
537 TAAACAAAAGCAAAAGCAAGCAACAAAAGTCAATCTGCTGTAATCTCAAA 596
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
336 TACTGATTAAGAAAAGCAAGCAAGTCAATGATGATGATGATGATGATGAT 395
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
597 AGATGATTAATTAATGTTTGAATACATGATGATGATGATGATGATGAT 642
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
396 AGCTCAGTAAAGTGAATGATGATGATGATGATGATGATGATGATGATGAT 441
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 9

US-09-710-279-4090
Sequence 4090, Application US/09710279
Patent No. 6703492
GENERAL INFORMATION:
APPLICANT: KIMBERLY, WILLIAM JOHN
TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
FILE REFERENCE: PUS480US
CURRENT APPLICATION NUMBER: US/09/710,279
CURRENT FILING DATE: 2000-11-09
PRIOR APPLICATION NUMBER: 60/164,258
PRIOR FILING DATE: 1999-11-09
NUMBER OF SEQ ID NOS: 4472
SOFTWARE: Patent Ver. 2.1
SEQ ID NO 4090
LENGTH: 4642
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic
OTHER INFORMATION: nucleic acid sequence
US-09-710-279-4090

Query Match 3.4%; Score 39.6; DB 3; Length 4642;
Best Local Similarity 43.6%; Pred. No. 0.96;
Matches 177; Conservative 0; Mismatches 229; Indels 0; Gaps 0;

```
237 AAAGCGATTAAACGGAACCTCAGGGCCGAATATCGAGACGAAAGATCATAG 296
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2052 AATGCAATTAAAGATGAATTAATGTTTGTAAGATTTGCTTTAGCATTCAGAG 2111
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
297 AGAAACATGAAAGAGGGCGTTACCGCATTTGTACGTAATGCAATGGCTTTAC 356
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2112 AATTCAGATGAGAGGCTTTGAAAGTAGTTCAGTTAACAATTACAGATGACAT 2171
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

QY 357 TTATCAATCATCTTTCAGAGAACTTCTGAGGCCAATATGCGCGCTATTGGAATGT 416
Db 2172 GTTAGCTCATTTATTAATAATGATTAATCAAGAGTCTTCTGAGAAAGTTGAATG 2231
QY 417 TTTGAAGACCATATGATGTCATGCGGCCGACGCTTGATGAAATTCGAGAAAGCTTTAC 476
Db 2232 TATGGAAGGTGATTCGCTGTGAAACGTAAGAAATTAATCAATTCGATGAACAGATGC 2291
QY 477 CGTACAAATTCCTTATTAATGCGCATCTTGTCAATACAGATGTAATGATCCGAGTTCTT 536
Db 2292 TGGTAATATTCACATGCGGCGATTTAGATATGACGTATATTAAGATGATCTGTTCTA 2351
QY 537 TAAACAAAAGCAAGAAAGCAACAAAGTCATCATGCTGATACCTCAAAAATTCAG 596
Db 2352 TACTGATTAAGAAAGAAAGCAAGCTCACATGATGACAGTCTTAAGAAAGATTAATCTC 2411
QY 597 AGATGATATTTACGTAATTTTGAATACATGATGTAATCCCTGATAC 642
Db 2412 AGCTCCAGCTAAAGTATGATTAAGAAACATGTAATTCACACTAAC 2457

RESULT 10
5231168-1
; Patent No. 5231168
; APPLICANT: DZIEGIEL, MORTEN,BORE, MARTIN,JEPSSEN, SOREN;
; VUUST, JENS,RIENECK, KLAUS,WIND, ANNETTE,JAKOBSEN, PALLE H.
; TITLE OF INVENTION: MALARIA ANTIGEN
; NUMBER OF SEQUENCES: 19
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/409,658
; FILING DATE: 18-SEP-1989
; SEQ ID NO:1;
; LENGTH: 3095
5231168-1

Query Match 3.3%; Score 39.2; DB 9; Length 3095;
Best Local Similarity 46.2%; Pred. No. 1;
Matches 168; Conservative 0; Mismatches 193; Indels 3; Gaps 1;

QY 281 AAAAAAGAGTCATGAGAGAAACAGTAGAAAGAGGGCTAACGCGATTGTCAAGTAATGCA 340
Db 878 AAGAAAAACAAATTCACAGAAATCAGTTGAAGAAATTCAGTAATGAGATGAAATTTG 937
QY 341 TGGCTGTTAACCCAGATATTAATCAATCTTTCAGAGAAACTTGTGAGCGCAATATCG 400
Db 938 AAGATGTTCAACACTGAAACATTTAGATTCATMAAAACAGTTGATCCAGAAATAGTAG 997
QY 401 GCGTCATTGTGAATGTTTGAAGACCATATGATGTCATGG--GGCCGACGCTTGATG 457
Db 998 AAGTTGAAGAAATTCCTTCAGAACTACATGAAATGAAATGAGTGGCTCATCCAGAAATTTGTTG 1057
QY 458 AAATTGCAAGAGCGTTTACCGCTACCAATTCCTTATTAATGGCCATGTTGTCAATTACAGATA 517
Db 1058 AAATTGAGGAAGTTTCTCTGAACCAATCAAAAATTAAGAAATTTCAATAATTAATGAAG 1117
QY 518 GTGAATATACCGAGTCTTTTAAACAAAAGCAAAAGCAAAACCAAAAGTCATCATTTG 577
Db 1118 ATGATTAAGATGCACATATTCAGCATGAAATGATGAGATGAAAGAAATTAATCTTCAGAAAG 1177
QY 578 CTGATTAACCAAAAATTAACAGATGATGATTTTACGTAATTTTGAATACATGATATCCCTG 637
Db 1178 ATGATTAAGAAATGAAAGATTTGAACATGAAATGATGAAAGTTTGAAGAAATTTCAACAGAAAG 1237
QY 638 ATAA 641
Db 1238 ATAA 1241

RESULT 11
US-09-710-279-3209
; Sequence 3209, Application US/09710279
; Patent No. 6703492

; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: P31480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; PRIOR FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3209
; LENGTH: 1011
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-09-710-279-3209

Query Match 3.3%; Score 38.6; DB 3; Length 1011;
Best Local Similarity 43.5%; Pred. No. 0.92;
Matches 176; Conservative 0; Mismatches 229; Indels 0; Gaps 0;

QY 238 AAGCCGATTAAGCGAAACCTCAGGGCGCAATATCGAGAGCAAAAAGATCATGAGA 297
Db 1 ATGCGCAATTAAGATGCAATTAATGTTTGTGTAAGATTTGTTAGCATTCAGAGA 60
QY 298 GAAACATGAAAGAGGGCTTAACCGCATTTGTCAAGTAATGCAATGCGTGTAAACCCAGAT 357
Db 61 ATTCAATATGTAAGAGTCTTGAAGTAGTTCAGATTAAACACTTAACAGATGACGATATG 120
QY 358 TATCAATATCATCTTTCAGAAAGAACTTCGACAGCCAAATATCGCGCTCATTTGTAATGTT 417
Db 121 TTAGCTCATTTATTAATAATACATATCATATGCAATGATGATGATGATGATGATGATGATG 180
QY 418 TTGAAGACCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 477
Db 181 ATCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 240
QY 478 GCTACATTTCTTATTAATGCGCATCTTGTCAATACAGATGATGATGATGATGATGATGATG 537
Db 241 GGTAAATTAACATGAGGGCGGATTTAGATATGACGATGATGATGATGATGATGATGATGATG 300
QY 538 AAACAAAGCAAGAAAGCAACAAACAAAGTCATCTGATGATGATGATGATGATGATGATGATG 597
Db 301 ACTGATTAAGAAAGCAACAAAGTCATGATGATGATGATGATGATGATGATGATGATGATGATG 360
QY 598 GATGATATTTACGTAATTTTGAATATACATGATGATGATGATGATGATGATGATGATGATG 642
Db 361 GCTCAGCTAAAGGTGATGTAAGAAACATGATGATGATGATGATGATGATGATGATGATGATG 405

RESULT 12
US-09-134-000C-2669
; Sequence 2669, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2669
; LENGTH: 1368
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-134-000C-2669

Query Match 3.2%; Score 38.4; DB 3; Length 1368;
Best Local Similarity 48.2%; Pred. No. 1.2; Indels 116; Gaps 0;
Matches 108; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 391 GCCATATCGGCGCATTTGTGATGTTTGAAGACCATATGATGATGAGGCGGACG 450
DB 736 GCGCTGAAGAAGTCCCTTCGAAATTCAGAAAGAAAGTATGTCTATTCGACTGA 795
QY 451 CTGTGTAATTCGAAAGCGTTTACCGTACATTCCTTATATGCGCATCTTGTACT 510
DB 796 AATCTGTGAAGCAAGAGTTAGTTCAATTTTGGAAAAAGATATACAGTTAAG 855
QY 511 ACAGATAGTATATACGAGTCTTTTAAACAAAAGCAAGAAAGCAACAAAGTC 570
DB 856 ACGTATATAGATATTCGCAATTCGCAAGTGTAGAAAGAACTGTGCAATTTGAAGA 915
QY 571 ATCATTTGCTGATACCTCAAAAATTCAGATGAGTATTTACGTAA 614
DB 916 AATGCTCGTTAAAGCTGAAACAAATTCGCAAAATTTTACAGAA 959

RESULT 13
US-09-134-001C-2329
Sequence 2329 Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
PRIOR FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 2329
LENGTH: 1158
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2329

Query Match 3.2%; Score 38.2; DB 3; Length 1158;
Best Local Similarity 50.3%; Pred. No. 1.3;
Matches 94; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 452 TTGATGAATTCGAGAGCGTTTACCGCTACAAATTCCTATATATGCCCATCTTGCTATTA 511
DB 224 TTGTTGAAATTCAGAAACAGATCAGTTGTACTTCTGTCCTTCTTCTGTTGATATTA 283
QY 512 CAGATGATGATATACCGAGTTCTTTAAACAAAAGCAAGAAAGCAACAAAGTGA 571
DB 284 TTAATAAATCTCTCGAAAAAGATTAATTAATCACTAATGAAATTCGAAAGCTTAA 343
QY 572 TCATTGCTGATACCTCAAAAATTCAGATGAGTATTTAGCTAATTTTGAATACATGAT 631
DB 344 TCACATCAGTCACTTCAAGATTTAAGTGCGTTAAGTCTGATCATATACCATTAAT 403
QY 632 TCCCTGA 638
DB 404 TACCTGA 410

RESULT 14
US-08-323-170B-1
Sequence 1, Application US/08323170B
Patent No. 5733772
GENERAL INFORMATION:
APPLICANT: Williamson, Kim C.
APPLICANT: Kaslow, David C.
TITLE OF INVENTION: Cloning and Expression of Plasmodium
TITLE OF INVENTION: falciiparum Transmission-Blocking Target Antigen, Pf230

NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/323,170B
FILING DATE: 13-OCT-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/010,409
FILING DATE: 29-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Quine, Jonathan A.
REGISTRATION NUMBER: P-41,261
REFERENCE/DOCKET NUMBER: 015280-113100US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 9636 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 149..9556
US-08-323-170B-1

Query Match 3.2%; Score 38.2; DB 2; Length 9636;
Best Local Similarity 55.7%; Pred. No. 3.4;
Matches 73; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 921 AATGACGATTCCTTATATGAGCAAGTGAATCTTAATCGGTGAACACAGA 980
DB 7309 AATTTCTTTTAACTTAATATGAAAAATATGCACTACTATTAGATGAAAGAAAGCA 7368
QY 981 ACCGATCGTAAAGCCTATGAGCAAGCAAAATTCCTGAGCAAACTGCATGACCTAGA 1040
DB 7369 AAAAAAATTAACACAAACAGAAAGAAAGCAACAAATTTTAAAGATCAAGA 7428
QY 1041 GTATAGTCA 1051
DB 7429 TGATGATTTA 7439

RESULT 15
US-08-954-441-1
Sequence 1, Application US/08954441
Patent No. 6316000
GENERAL INFORMATION:
APPLICANT: Williamson, Kim C.
APPLICANT: Kaslow, David C.
TITLE OF INVENTION: Cloning and Expression of Plasmodium
TITLE OF INVENTION: falciiparum Transmission-Blocking Target Antigen, Pf230
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/954,441
FILING DATE: 20-OCT-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/323,170
FILING DATE: 13-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/010,409
FILING DATE: 29-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Einhorn, Gregory P.
REGISTRATION NUMBER: 38,440
REFERENCE/DOCKET NUMBER: 015280-113110US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 9636 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 149..9556
US-08-954-441-1

Query Match 3.2%; Score 38.2; DB 3; Length 9636;

Best Local Similarity 55.7%; Pred.No.3.4;

Matches 73; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY	921	AAATGACGTATTGCTTATTTATGAAGCAAGTGAAGTCTTATCGTGAAACACACAGA	980
DB	7309	AAATTTCTTTACTTATATAGAAAAATATGCACTACTATTAGATGAGAAAGACA	7368
QY	981	ACCGATCGTAAAGCCTATGAAGAGCAAAATTCCTGCAGCAAACTGCATGACCTAGA	1040
DB	7369	AAAAAAATTAACACACACAGAGAGAACACACAAATTTAAAGATCAAGA	7428
QY	1041	GTATAGTCAA	1051
DB	7429	TGATGATTAA	7439

Search completed: February 27, 2006, 06:46:09
Job time : 287.342 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:33:05 ; Search time 2573.57 Seconds
(without alignments)
9873.061 Million cell updates/sec

Title: US-10-789-164-2.

Perfect score: 447
Sequence: 1 atgttcgacgacgattata.....tctttgttactactaact 447

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl :
1: gb_da :
2: gb_in :
3: gb_env :
4: gb_cm :
5: gb_ov :
6: gb_pac :
7: gb_ph :
8: gb_pr :
9: gb_ro :
10: gb_str :
11: gb_sy :
12: gb_un :
13: gb_vl :
14: gb_hcg :
15: gb_pl :

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	445.4	99.6	2775	1 DQ086153	DQ086153 Bacillus
2	443.8	99.3	4315	1 AB039950	AB039950 Bacillus
3	442.2	98.9	2989	1 AB016245	AB016245 Bacillus
4	442.2	98.9	8189	1 BS292954	BS292954 B. subtilis
5	442.2	98.9	200690	1 BSUB0019	BSUB0019 Bacillus su
6	440.6	98.6	450	6 ES0426	ES0426 Process for
7	440.6	98.6	3045	6 ES0426	ES0426 Process for
8	435.8	97.5	3738	1 AB046355	AB046355 Bacillus
9	290.8	65.1	110000	1 CP000002_36	Continuation (37 o
10	280.8	65.1	110000	6 AX434510	AX434510 Sequence
11	289.2	64.7	450	6 BACCAPAC	BACCAPAC
12	237.4	53.1	3244	1 AB011191	AB011191 Bacillus
13	237.4	53.1	94829	1 AB011335	AB011335 Bacillus
14	237.4	53.1	96231	1 AF188935	AF188935 Bacillus
15	237.4	50.1	930	1 AB125962	AB125962 Bacillus
16	223.8	44.6	1336	1 BF060883	BF060883 Bacillus fi
17	199.4	44.5	110000	1 AP006716_05	Continuation (6 of
18	199				

c 19	191	42.7	300275	1 AE016751	AE016751 Staphyloc
c 20	189.4	42.4	453	6 AR484743	AR484743 Sequence
c 21	189.4	42.4	453	6 AX143601	AX143601 Sequence
c 22	189.4	42.4	110000	1 CP000029_21	Continuation (22 o
c 23	182	40.7	3240	1 AF269703	AF269703 Staphyloc
c 24	182	40.7	3240	1 AR485657	AR485657 Sequence
c 25	182	40.7	3240	6 AX145021	AX145021 Sequence
c 26	180.4	40.4	110000	1 AP008934_03	Continuation (4 of
c 27	118.8	26.6	110000	1 BA000028_02	Continuation (3 of
c 28	53.8	12.0	110000	1 AE009951_15	Continuation (16 o
c 29	45.6	10.2	3029	1 AF270157	AF270157 Staphyloc
c 30	45.6	10.2	3029	6 AR486111	AR486111 Sequence
c 31	45.6	10.2	3029	6 AX145475	AX145475 Sequence
c 32	45.6	10.2	110000	1 AJ749949_08	Continuation (9 of
c 33	43.2	9.7	158981	14 AC016242	AC016242 Homo sapi
c 34	43.2	9.7	162151	8 AC012358	AC012358 Homo sapi
c 35	41.2	9.2	110520	15 AC148304	AC148304 Medicago
c 36	41.2	9.2	125303	15 AC147009	AC147009 Medicago
c 37	41.2	9.2	127018	15 AC135414	AC135414 Medicago
c 38	39.8	8.9	15319	15 AB023035	AB023035 Arabidops
c 39	39.6	8.9	221912	9 AL672003	AL672003 Mouse DNA
c 40	39.6	8.9	349980	6 AX344554	AX344554 Sequence
c 41	39.4	8.8	160286	15 AY780259	AY780259 Eucalyptu
c 42	39.4	8.8	175820	8 AL162378	AL162378 Human DNA
c 43	39.2	8.8	86901	9 AC105630	AC105630 Rattus no
c 44	39.2	8.8	200289	9 AC122295	AC122295 Mus muscu
c 45	39.2	8.8	232930	9 AC113761	AC113761 Rattus no

ALIGNMENTS

RESULT 1
DQ086153 2775 bp DNA linear BCT 27-JUN-2005
LOCUS Bacillus subtilis strain ZJU-7 Pgab (pgab), Pgac (pgac), and Pgaa
DEFINITION (pgaa) genes, complete cds.
ACCESSION DQ086153
VERSION DQ086153.1 GI:68138277

KEYWORDS Bacillus subtilis
SOURCE Bacillus subtilis
ORGANISM Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

REFERENCE 1 (bases 1 to 2775)
Shi, F., Xu, Z., and Cen, P.
Efficient production of poly(glutamic acid) by a new strain
Bacillus subtilis ZJU-7

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2775)
Shi, F., Xu, Z., and Cen, P.
Direct Substitution
Submitted (04-JUN-2005) Institute of Bioengineering, Department of
Chemical Engineering and Bioengineering, Zhejiang University, No.38,
Zheda Road, Hangzhou, Zhejiang 310027, P.R. China

JOURNAL

FEATURES

source
1..2775
/organism="Bacillus subtilis"
/mol_type="genomic DNA"
/strain="ZJU-7"
/db_xref="taxon:1423"
1..1182
/gene="pgab"
1..1182
/gene="pgab"
/function="Biosynthesis of poly(glutamic acid)"
/codon_start=1
/product="Pgab"
/protein_id="AY85940.1"
/translation="MWLLIACAVILVIGLEKRRHQKIDALPVAVNINIGRKSTV
TRLTGILIEAGYKTVGKTGDARMTVMDPEEKIKRKPQPNIGGEKVEWRTVE
RGNALVSECMVNPYQIIIFQELLQANIGVTAVNLIEDHMDVGPFLDEIAEFTAT

IPYNGHIVITDSEYTEFEKOKAKERNTKVIILADNSKIDEXYLKREYVW/PDNLAL
GVAOLGIDEETAFKGMNAPDPGAMILPLISSEPHVNGFPAANDASTLINIK
RKEIGYPTDPIIIMNCRADRVDRTOOFANDVLPYIEASELILIGETTERIVAYEE
GKIPADKLDHELYKSTDEIMELKKRMNRVIYGVNIGHGAEPLEIKIHEYKQOLV
S"

gene
CDS

1183. 1632
/gene="pgac"
1183. 1632
/gene="pgac"
/function="biosynthesis of poly(glutamic acid)"
/codon_start=1
/transl_table=11
/product="Pgac"
/protein_id="AAV85941.1"
/db_xref="GI:68138279"
/translation="MFGSDIYALITLILFAEKTGIVPAGLVVPGYLVGNOP
VFILVLVSLTIVYIVKYGSKFMILYGRKRFAMLTGTVLAKTAPFLYIPFEI
AEFRIGIIVPELANTIOKOGLTTFGSLILSGATPALMFVYLLI"
1633. 2775
/gene="pgaa"
1633. 2775
/gene="pgaa"
/function="biosynthesis of poly(glutamic acid)"
/codon_start=1
/transl_table=11
/product="Pgaa"
/protein_id="AAV85942.1"
/db_xref="GI:68138280"
/translation="MKKELSFHEKLLKLTQOKKKTKNHVFIAPIVFLMFAFMAG
KAETPKVTSDDVLSASFVGDIMMGRVEKTEKGDASIPOVEPIFRASDVAGN
FENPVYOKNYKQADKEIHLQTNKESVYVLDKMNFTVANSANNHAMDYGVGMKDTLG
EFAKONLDIVGAGYSLDAKKISYQVNGVTIATLTFDVSIGKFAKNTPGVLP
DPEIFIMISEAKKHADIWVQSHWGGEYNDPDROROLARASDAGADIIVGHHPH
VLBEIYVNGTVIFYSLGNFVFDGWTTRDSALVQVHLKKNGTGREVTPIDIHET
PAPVKDSLKQKTIIRLTKDSNFAKVEDKLTVIDHSDLSK"

gene
CDS

1 ATGTTCCGATCGAATTATATACATCGCACTAATTTTAAAGTGTACTACTAGTTAAATTTT 60
1183 ATGTTCCGATCGAATTATATACATCGCACTAATTTTAAAGTGTACTACTAGTTAAATTTT 1242

ORIGIN

Query Match 99.6%; Score 445.4; DB 1; Length 2775;
Best Local Similarity 99.8%; Pred. No. 98-105;
Matches 446; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGTTCCGATCGAATTATATACATCGCACTAATTTTAAAGTGTACTACTAGTTAAATTTT 60
DB 1183 ATGTTCCGATCGAATTATATACATCGCACTAATTTTAAAGTGTACTACTAGTTAAATTTT 1242
QY 61 GGGGAAAAACAGGAGATCGTCCGCGACAGACTTGTGTACCCGGATATTAGACCTTGTG 120
DB 1243 GCGGAAAAACAGGAGATCGTCCGCGACAGACTTGTGTACCCGGATATTAGACCTTGTG 1302
QY 121 TTAAATCAGCCGCTCTTATTTTACTGTTTGTAGTACGAGCTGCTCACTATGTTATC 180
DB 1303 TTAAATCAGCCGCTCTTATTTTACTGTTTGTAGTACGAGCTGCTCACTATGTTATC 1362
QY 181 GTGAAATACGCTTTATCCAAATTTATGATTTTGTACGACGACGAAAAATTCGCTCCATG 240
DB 1363 GTGAAATACGCTTTATCCAAATTTATGATTTTGTACGACGACGAAAAATTCGCTCCATG 1422
QY 241 CTGATTAACAGGAGATCGTCTTAAATTCGCTTGTATTTCTATACCCGATTTGTACATTT 300
DB 1423 CTGATTAACAGGAGATCGTCTTAAATTCGCTTGTATTTCTATACCCGATTTGTACATTT 1482
QY 301 GAAATCGCGAATTTTCGAGGAATCGGACATCGAGCCAGGTTAATGCGCAATACCAAT 360
DB 1483 GAAATCGCGAATTTTCGAGGAATCGGACATCGAGCCAGGTTAATGCGCAATACCAAT 1542
QY 361 CAGAAACAGGTTTAAACATTAAGTTTCGAGACAGCTGCTATTTAGCGGACGACCTTT 420
DB 1543 CAGAAACAGGTTTAAACATTAAGTTTCGAGACAGCTGCTATTTAGCGGACGACCTTT 1602
QY 421 GCTATCATGTTTGTATTACTTAAAT 447
DB 1603 GCTATCATGTTTGTATTACTTAAAT 1629

RESULT 2
AB039950
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AB039950 4315 bp DNA linear BCT 12-SEP-2000
Bacillus subtilis capB, capC, capA, ywec genes, complete cds.
AB039950
AB039950.1 GI:10119860
CAPA; CapC; CapB.
Bacillus subtilis
Bacillus subtilis
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
1 (bases 1 to 4315)
Tran, L.P. and Itoh, Y.
Nucleotide sequence of the capBCA operon in Bacillus subtilis
(nato)

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

2 (bases 1 to 4315)
Tran, L.P. and Itoh, Y.
Direct Submission
Submitted (08-MAR-2000) Yoshifumi Itoh, National Food Research
Institute, Applied Microbiology; Kannondai 2-1-2, Tsukuba, Ibaraki
305-8642, Japan (E-mail: yosifumi@nfri.affrc.go.jp,
Tel: +81-298-38-8075, Fax: +81-298-38-7956)

FEATURES
source

1. 4315
Location/Qualifiers

/organism="Bacillus subtilis"
/mol_type="genomic DNA"
/strain="Asehi-kawa"
/db_xref="taxon:1423"
941. 2122
/gene="capB"
941. 2122
/gene="capB"
/codon_start=1
/transl_table=11
/product="CapB"
/protein_id="BAB13484.1"
/db_xref="GI:10119861"
/translation="MMLIIICAVILVIGLEKRRHQNIDALPVRVNINGIGKSTV
TRLTGILIRAGYKVTGKTGTGTABMYMTPEPKPIKRPQGNIGDEKEMRETV
RGAAATVSECAVNPVQIIFQEBELQANIVINVEDIMDWDGPIRDIARFAT
IPYNGHIVITDSEYTERFEKOKAKERNTKVIILADNSKIDEXYLKREYVW/PDNLAL
GVAOLGIDEETAFKGMNAPDPGAMILPLISSEPHVNGFPAANDASTLINIK
RKEIGYPTDPIIIMNCRADRVDRTOOFANDVLPYIEASELILIGETTERIVAYEE
GKIPADKLDHELYKSTDEIMELKKRMNRVIYGVNIGHGAEPLEIKIHEYKQOLV
S"

2137. 2586
/gene="capC"
2137. 2586
/gene="capC"
/codon_start=1
/transl_table=11
/product="CapC"
/protein_id="BAB13485.1"
/db_xref="GI:10119862"
/translation="MFGSDIYALITLILFAEKTGIVPAGLVVPGYLVGNOP
VFILVLVSLTIVYIVKYGSKFMILYGRKRFAMLTGTVLAKTAPFLYIPFEI
AEFRIGIIVPELANTIOKOGLTTFGSLILSGATPALMFVYLLI"
2605. 3747
/gene="capA"
2605. 3747
/gene="capA"
/codon_start=1
/transl_table=11
/product="CapA"
/protein_id="BAB13486.1"
/db_xref="GI:10119863"
/translation="MKKELSFHEKLLKLTQOKKKTKNHVFIAPIVFLMFAFMAG
KAETPKVTSDDVLSASFVGDIMMGRVEKTEKGDASIPOVEPIFRASDVAGN
FENPVYOKNYKQADKEIHLQTNKESVYVLDKMNFTVANSANNHAMDYGVGMKDTLG
EFAKONLDIVGAGYSLDAKKISYQVNGVTIATLTFDVSIGKFAKNTPGVLP
DPEIFIMISEAKKHADIWVQSHWGGEYNDPDROROLARASDAGADIIVGHHPH
VLBEIYVNGTVIFYSLGNFVFDGWTTRDSALVQVHLKKNGTGREVTPIDIHET
PAPVKDSLKQKTIIRLTKDSNFAKVEDKLTVIDHSDLSK"

gene
CDS

2137. 2586
/gene="capC"
2137. 2586
/gene="capC"
/codon_start=1
/transl_table=11
/product="CapC"
/protein_id="BAB13485.1"
/db_xref="GI:10119862"
/translation="MFGSDIYALITLILFAEKTGIVPAGLVVPGYLVGNOP
VFILVLVSLTIVYIVKYGSKFMILYGRKRFAMLTGTVLAKTAPFLYIPFEI
AEFRIGIIVPELANTIOKOGLTTFGSLILSGATPALMFVYLLI"
2605. 3747
/gene="capA"
2605. 3747
/gene="capA"
/codon_start=1
/transl_table=11
/product="CapA"
/protein_id="BAB13486.1"
/db_xref="GI:10119863"
/translation="MKKELSFHEKLLKLTQOKKKTKNHVFIAPIVFLMFAFMAG
KAETPKVTSDDVLSASFVGDIMMGRVEKTEKGDASIPOVEPIFRASDVAGN
FENPVYOKNYKQADKEIHLQTNKESVYVLDKMNFTVANSANNHAMDYGVGMKDTLG
EFAKONLDIVGAGYSLDAKKISYQVNGVTIATLTFDVSIGKFAKNTPGVLP
DPEIFIMISEAKKHADIWVQSHWGGEYNDPDROROLARASDAGADIIVGHHPH
VLBEIYVNGTVIFYSLGNFVFDGWTTRDSALVQVHLKKNGTGREVTPIDIHET
PAPVKDSLKQKTIIRLTKDSNFAKVEDKLTVIDHSDLSK"

gene
3762..3929
/gene="ywcC"
/gene="ywcC"
/gene="ywcC"
/codon_start=1
/transl_table=11
/protein_id="BAB13487.1"
/db_xref="GI:10119864"
/translation="MKFVAIWPFAVAIVFMFMSAFKENDQLTDOEKOKIDMENMKI
QOQEEPVANK"

ORIGIN

Query Match 99.3%; Score 443.8; DB 1; Length 4315;
Best Local Similarity 99.6%; Pred. No. 2.2e-104; Mismatches 2; Indels 0; Gaps 0;
Matches 445; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGTTGGATTCAGATTATATACATCGCATTAATTTAGTGTACTACTAGTTAATTTT 60
DB 2137 ATGTTGGATTCAGATTATATACATCGCATTAATTTAGTGTACTACTAGTTAATTTT 2196
QY 61 GCGGAAAAACAGAGATCGTCCGCGAGACCTTGTACCGGATATTTAGACCTGTG 120
DB 2197 GCGGAAAAACAGAGATCGTCCGCGAGACCTTGTACCGGATATTTAGACCTGTG 2256
QY 121 TTTAATCAGCCGCTTTATTTATTTCTGTTGCTAGTGAAGCTTCTCACTATATGTTATC 180
DB 2257 TTTAATCAGCCGCTTTATTTATTTCTGTTGCTAGTGAAGCTTCTCACTATATGTTATC 2316
QY 181 GTGAATACGGCTTATCCAAATTTATGATTTTGTACGAGCGAGAAATTCGTGCATG 240
DB 2317 GTGAATACGGCTTATCCAAATTTATGATTTTGTACGAGCGAGAAATTCGTGCATG 2376
QY 241 CTGATACAGGATCGTCTTAAATTCGCTTGTATTTCTATACCGATTTGACCATTT 300
DB 2377 CTGATACAGGATCGTCTTAAATTCGCTTGTATTTCTATACCGATTTGACCATTT 2436
QY 301 GAAATCGCAAAATTCGAGGAATCGGCATCATGTGCGAGCTTAAATTCGCAATACCAT 360
DB 2437 GAAATCGCAAAATTCGAGGAATCGGCATCATGTGCGAGCTTAAATTCGCAATACCAT 2496
QY 361 CAGAAACAGGTTAACCATTAAGTTCGAGACACGCTGATTTGACGAGGACCTTT 420
DB 2497 CAGAAACAGGTTAACCATTAAGTTCGAGACACGCTGATTTGACGAGGACCTTT 2556
QY 421 GCTATCATGTTTGTACTACTTAATTT 447
DB 2557 GCTATCATGTTTGTACTACTTAATTT 2583

RESULT 3
AB016245 2989 bp DNA linear BCT 14-OCT-1999

LOCUS Bacillus subtilis pgsA, pgsB and pgsC genes, complete cds.
DEFINITION
AB016245
AB016245.1 GI:6045071
VERSION
KEYWORDS
SOURCE
ORGANISM
Bacillus subtilis
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
Ashichu, M., Soda, K. and Misono, H.
A poly-gamma-glutamate synthetic system of Bacillus subtilis IFO
3336: gene cloning and biochemical analysis of poly-gamma-glutamate
produced by Escherichia coli clone cells
Biochem. Biophys. Res. Commun. 263 (1), 6-12 (1999)

REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
Ashichu, M.
Direct Submission
Submitted (14-JUL-1998) Makoto Ashichu, Kochi University, Research
Institute of Molecular Genetics, Otsu 200, Monobe, Nankoku, Kochi
783-8502, Japan (E-mail: ashichu@img.kochi-u.ac.jp)

FEATURES
source
Tel: 0888-64-5215, Fax: 0888-64-5109

1..2989
/organism="Bacillus subtilis"
/mol_type="genomic DNA"
/strain="IFO 3336"
/sub_species="Bacillus natto"
/db_xref="taxon:1423"
1..1182
/gene="pgsB"
1..1182
/gene="pgsB"
/codon_start=1
/transl_table=11
/product="pgsB"
/protein_id="BA85263.1"
/db_xref="GI:6045072"
/translation="MMLLIACAVILVIGLEKRRHQNIDALPVNVINIGIKGSTV
TRLTGLILEAGYKTVGKTGTGDAMIYMDPEEKPIKRPQSPISIQEKVRETFE
RGAAIVSECAVNDYOIIFOEILLQNIYIVVLEDDHMDVMPDLEIAEATAT
IPYNGHIVITDSEVPEFKOKAKERNVTIADNSKIDEXYLKREYVFPDNASIAL
GVQMLGIDETAFKGLNAPPDGAMTILPISREGEVNGFAANDASTNIAWK
RKVEIGYPLDDPIIIMNCRAKRVDTQOPADVLPYIASLLILGETTEPIVAYEE
GKIPADKLHDLKSTDEIMELKKSMHNRVIYGVNIGHAEPLEIKHEKVAOLV
S"

gene
1197..1646
/gene="pgsC"
1197..1646
/gene="pgsC"
/codon_start=1
/transl_table=11
/product="pgsC"
/protein_id="BA85264.1"
/db_xref="GI:6045073"
/translation="MFGSDLYIALILGVLSLIFAEKTIQVPAQLVBEYGLVFNOP
VEILVLVSLTYIVIVKYGSLKFMILYGRKFAAMLITGIYKIAFPLFIVPEI
AEFRGIIIVPLIANTIOKQGLTTFSTLLLSGATPAIMFVYLI"

gene
1665..2807
/gene="pgsA"
1665..2807
/gene="pgsA"
/codon_start=1
/transl_table=11
/product="pgsA"
/protein_id="BA85265.1"
/db_xref="GI:6045074"
/translation="MKKELSFHEKLKLTQKQKKTKNKHVIAPIVFLMFAFMAG
KAEPTVKTYSDDVLSASFVDIMMGRYVEKTEBQADSIQYVEPIRADYVAGN
FENPVYOKNRYQADKEIHLQTNKESVVKLKMNFVLSANNHAMDYVGQMKDTLG
EBAKQNDIVGAGYSLSDAKKISYQKNGVITATLGFDPVSGKPAADKNTIVGHHP
DPEIFLPMISEAKKHADIVVDSHNGOEYDNDPNRQRLAAMDAGADKIIYGHHP
VLEPIEVNGLVIFYSLGNFVDPQWTRDRDALVOYHLKNGTGFEVTPPIDHEAT
PAPVKDLSIKOKTIIRELTKDSNFAKVEDGKLTDPIDHSDLSKSK"

ORIGIN

Query Match 98.9%; Score 442.2; DB 1; Length 2989;
Best Local Similarity 99.3%; Pred. No. 6e-104; 3; Indels 0; Gaps 0;
Matches 444; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGTTGGATTCAGATTATATACATCGCATTAATTTAGTGTACTACTAGTTAATTTT 60
DB 1197 ATGTTGGATTCAGATTATATACATCGCATTAATTTAGTGTACTACTAGTTAATTTT 1256
QY 61 GCGGAAAAACAGAGATCGTCCGCGAGACCTTGTACCGGATATTTAGACCTGTG 120
DB 1257 GCGGAAAAACAGAGATCGTCCGCGAGACCTTGTACCGGATATTTAGACCTGTG 1316
QY 121 TTTAATCAGCCGCTTTATTTATTTCTGTTGCTAGTGAAGCTTCTCACTATATGTTATC 180
DB 1317 TTTAATCAGCCGCTTTATTTATTTCTGTTGCTAGTGAAGCTTCTCACTATATGTTATC 1376
QY 181 GTGAATACGGCTTATCCAAATTTATGATTTTGTACGAGCGAGAAATTCGTGCATG 240

Db	1377	GTGAATACGGTTTATCCAAATTATGATTTGTACGAGCAGAAAATTCGCTGCCATG	1436	gene	join(1069..1075,1084..1533) /gene="ywcA"
Qy	241	CTGATTAACAGGATGCTCTTAATAATCGCGTTTGATTTTCTATACCCGATTTGTACCATTT	300	RBS	/gene="ywcA" 1069..1075
Db	1437	CTGATTAACAGGATGCTCTTAATAATCGCGTTTGATTTTCTATACCCGATTTGTACCATTT	1496	CDS	/gene="ywcA" 1084..1533
Qy	301	GAATCGCAGATTTGAGAGATCGGATTCATCGCAGGTTTAAATGGCAATACATTT	360		/note="product highly similar to Bacillus anthracis CapC protein"
Db	1497	GAATCGCAGATTTGAGAGATCGGATTCATCGCAGGTTTAAATGGCAATACATTT	1556		/codon_start=1 /transl_table=1 /protein_id="CAB07468.1"
Qy	361	CAGAAACAGGTTTAAACATTACGTTTGGAAACACAGCTGCTATTAGCGGACGACCTTT	420		/db_xref="GI:1894766"
Db	1557	CAGAAACAGGTTTAAACATTACGTTTGGAAACACAGCTGCTATTAGCGGACGACCTTT	1616		/db_xref="GOA:P96737"
Qy	421	GCTATCATGTTTGTACTACTAATTT	447		/db_xref="UniProt/TREMBL:P96737"
Db	1617	GCTATCATGTTTGTACTACTAATTT	1643	gene	/translation="MEGSDLYIALILGVLLSLFAEKRTGIVPAGLVVGYGLVFNOP VFILVIVLSLLVIVLYKGLSKEMILYGRKRAAMITGIVLKIAEDPIPIVPEPI AERFGIIVPGLIANTTOKOGLITTFGSTLLLSGATFAIMFYILI"
RESULT 4					1538..2694
LOCUS	BS292954	8189 bp	DNA	RBS	/gene="ywcB"
DEFINITION	B. subtilis ywc [A,B,C,D,E,F,G] and gerBC genes.				1538..1544
ACCESSION	Z92954			CDS	/gene="ywcB"
VERSION	Z92954.1	GI:1894764			1552..2694
KEYWORDS	gerBC gene; ywsc gene; ywfa gene; ywcb gene; ywcc gene; ywcd gene; ywcf gene; ywcf gene; ywcf gene.				/gene="ywcB"
SOURCE	Bacillus subtilis				/note="product highly similar to Bacillus anthracis CapA protein"
ORGANISM	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.				/codon_start=1 /transl_table=1 /protein_id="CAB07469.1"
REFERENCE	1 (bases 1 to 8189)				/db_xref="GI:1894767"
AUTHORS	Presecan E., Moszer I., Bourcier L., Cruz Ramos H.C., de la Fuente V., Hillo M.F., Lelong C., Schleich S., Sekowska A., Song B.H., Villani G., Kunst F., Danchin A. and Glaeser P.				/db_xref="UniProt/TREMBL:P96738"
TITLE	The Bacillus subtilis genome from gerBC (311 degrees) to licoR (334 degrees)				/translation="MKKELSPHRLKLTQOKKTNKHVIAIPIVLMAFPMWAG KAETPKVYTSDDVLSAPVGDIMMGHYVEKTEQKADSIPOYVEPIFASDVAGN FENPVYTKYKADKEIHLQTKKESIKVLCNNFVLYNANNNAMGYQGMKDTLG EPKQNIIDVIGAGSISDAKKISYQKYNVTATLGTFTSGSGFPAKNTGVLBA DPELFIWISDAKHADIVVQSHMGQETNDPNDRQRLARMSDAGDIYGHNRH VLEIEVYNGTVIFYSLGNFVDFQWTRTDSALVOYHLKNGRGREVPIDHEAT PAPVKDLSKQITIRLITRDSNFMKVEDGKLTFDIDHSKLSK"
JOURNAL	Microbiology (Reading, Engl.) 143 (Pt 10), 3313-3328 (1997)			gene	2697..2914
PUBMED	9353933				/gene="ywcC"
REFERENCE	2 (bases 1 to 8189)			RBS	2697..2702
AUTHORS	Lejong C., Glaeser P., Presecan E. and Danchin A.				/gene="ywcC"
TITLE	Bacillus subtilis gerB downstream sequence			CDS	2709..2876
JOURNAL	Unpublished				/gene="ywcC"
REFERENCE	3 (bases 1 to 8189)				/codon_start=1 /transl_table=1 /product="unknown"
AUTHORS	Glaeser P.				/protein_id="CAB07470.1"
TITLE	Direct Submission				/db_xref="GI:1894768"
JOURNAL	Submitted (13-MAR-1997) Philippe Glaeser, Regulation de l'Expression Genetique, Institut Pasteur, 28 Rue du Dr Roux, Paris, 75724, FRANCE				/db_xref="UniProt/Swiss-Prot:P96739"
FEATURES					/translation="MKFVKAIPFVAVAIVFMMSAFKFNDOULTDOKOKIDMEMNKI OOOEPPNANK"
source	1..8189	Location/Qualifiers		terminator	2880..2914
	/organism="Bacillus subtilis"				/gene="ywcC"
	/mol_type="genomic DNA"			gene	3000..4258
	/strain="168"				/gene="ywcD"
	/db_xref="taxon:1423"			RBS	3000..3007
	1..1069				/gene="ywcD"
gene	/gene="ywsc"			CDS	3017..4258
CDS	<1..1069				/gene="ywcD"
	/note="product highly similar to Bacillus anthracis CapB protein"				/note="product similar to E. coli NUPC protein and to Listeria species p60-related proteins"
	/codon_start=2				/transl_table=1 /protein_id="CAB07476.1"
	/transl_table=1				/db_xref="GI:1894769"
	/db_xref="GI:1894765"				/db_xref="GOA:P96740"
	/db_xref="UniProt/TREMBL:Q798J2"				/db_xref="UniProt/Swiss-Prot:P96740"
	/translation="RGKSTVRLTGTGILGAGYKTVGKTGGDAMIYMDTPEEKPIK RKPOGNGEOKWRETFVERGANIVSECMANVPDYOITROEELLOANICIVNVLE DHMDVMTGLDIAAFTATPIYNGHUVITDSEYTFPKQAKKENTYVITADNSKIT DEYLFKRYWFPDNASIALGVQAQLIDESTAFGMNAPDPGAMKILPLISSEIP GHFVNGFANPASSTINIMKRYKEIGYPTDPIIIMNCRADRVDTQOFADVPEYE ASELLIGETTEPIVKAVEEGKIPADKLDLEYSKSTDIEMELKKSMNRVYGVGNI HGAAPFLIEKTHEYKQVLVS"				NILGVOYKGGEPTEKRGPDPSGLIOYVFSKADILHPSVANVOYKIGAVPENKPPD ILFKKSGSTGYVTHDALYIGDQOMHNSQSKVITITNKKSSYNSGTYIGARRIA DPAIADVPQVEAKYIGVYVFGSTSPSGDGLVQYVFOQALGITYLPRAEQOM AVGKRVAPQNIKPDVVYFSTNYKTGISHAGIYAGARFLOASRSEKVTISYLSYEDW KSKMTGIRRFNLIITPRENPVSEATLYGSEVPYKQGVLPETGFDAGFQVYVYORA

Scanlan,E., Schleich,S., Schroeter,R., Scoffone,F., Sekiguchi,J., Sekowska,A., Seror,S.J., Serror,P., Shin,B.S., Soldo,B., Sorokin,A., Tacconi,B., Takagi,T., Takahashi,H., Takemaru,K., Takeuchi,M., Tamakoshi,A., Tanaka,T., Terpeira,P., Tognoni,A., Tosato,V., Uchiyama,S., Vandenbol,M., Vannier,F., Vassart,A., Viari,A., Wambuit,R., Wedler,E., Wedler,H., Weltzenegger,T., Winters,P., Wipac,A., Yamamoto,H., Yamane,K., Yasumoto,K., Yata,K., Yoshida,K., Yoshikawa,H.F., Zumbstein,E., Yoshikawa,H. and Danchin,A.

The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*

Nature 390 (6657), 249-256 (1997)

2 (bases 1 to 200690)

Kunst,F., Ogasawara,N., Yoshikawa,H. and Danchin,A.

Direct Submission

Submitted (27-JUN-2003) I. Moszer, A. Danchin, Institut Pasteur, Genetique des Genomes Bacteriens, 28 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr, adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45 68 89 48

On Jul 7, 2003 this sequence version replaced gi:2636029.

This entry contains data from release R16.1 of the Subtilist database. Further data on gene annotation and detailed information about changes from previous releases can be found at <http://genolist.pasteur.fr/Subtilist/>.

location/Qualifiers

1. 200690

/organism="Bacillus subtilis subsp. subtilis str. 168"

/mol_type="genomic DNA"

/strain="168"

/db_xref="taxon:224308"

/complement(114..2987)

/gene="uvrA"

/locus_tag="BSU35160"

/complement(114..2987)

/gene="uvrA"

/locus_tag="BSU35160"

/function="excision of ultraviolet light-induced pyrimidine dimers in DNA"

/codon_start=1

/transl_table=11

/product="excinuclease ABC (subunit A)"

/protein_id="CAB15533.1"

/db_xref="GI:2636042"

/db_xref="GOA:O34863"

/db_xref="InterPro:IPR003439"

/db_xref="InterPro:IPR003593"

/db_xref="InterPro:IPR004602"

/db_xref="Subtilist:BG12697"

/db_xref="UniProt/Swiss-Prot:O34863"

/translation="MAMDIEVKGARAHNLKVIDITPRDDLVVVTGLSGSGKSLAFVTEIYAEQGRRYVESLSAYARQPLGQMDKPDVAIDGLSPAISIDQTKTSRNPRTVGTVEIADYRLRLYARVGKPCPEHGIEITQTEQWVDRLIEPERTKQLQVAPVSGRKGAYKVLQEIQRKQGVYRVIDGEAEASDIELEKNKKSIEVLVIDRIVAEVGAARLSDSLFETLRIGEGRWMDIVGEELMSEHACHPCGSGISGELEPRLFSNPSPGACPTCDLGMKLEVDADLVTPNDQSLKENVAPWPTISQVYRQLLEAVCTHYGIDMDVPRVDLPHQDLKVLVSGSDLLYFRFENDPGQIRGEBLOFEGVLRNIERYKETSDYIREQMEQYMSQKSCPTCKGKRLKEKALVLDGRHIGITELSVADALAFEDLTLSEKDMQJANLILREIYERLSFLDVGDLVLSRAAGTLSGEAGORIRLATQIGSRISGLVYLIDEPISGLHQRNDRLISALKMRDIGNTLIVEHEDETMAADYLDIDPGAGIHGGOVISAGTPEWVEDPNSLTGSLSGKKFPLPERKRPGRYIEIKGASENNLKKVNAKPLIGFTTAVTGVSGSGKTLVNIHLKALCKLHKAKKPGSHKEITGDLHDKVIDIDQAPIGRTPRSNPATYTTVPPDIDRIVPQNTBAKRYKKGRFSPNPKGCEACRGGDIKIKELMHFLPDVYVPCVCHGKRYNRRETLVYKXKSIQVLTVFEDALSFENIPKIKRLQTLVYDVGIVTLQGPATTLSGEAGORYLASLHKRSTGRLLYLIDEPTTGAVHDIDARLLVLOELVNDGTVIVIEHNLIKTADYIVLDGEGAGCGGTIVASGPERITEVESYTGRLKIVIRDRKTRMSLLAKKETATS"

/complement(2995..4980)

/gene="uvrB"

/locus_tag="BSU35170"

/complement(2995..4980)

/gene="uvrB"

/locus_tag="BSU35170"

/function="excision of ultraviolet light-induced pyrimidine dimers in DNA"

/note="alternate gene name: dlnA, uvrA"

/codon_start=1

/transl_table=11

/product="excinuclease ABC (subunit B)"

/protein_id="CAB15534.1"

/db_xref="GI:2636043"

/db_xref="GOA:P37954"

/db_xref="InterPro:IPR001410"

/db_xref="InterPro:IPR001650"

/db_xref="InterPro:IPR001943"

/db_xref="InterPro:IPR004807"

/db_xref="InterPro:IPR006935"

/db_xref="InterPro:IPR009055"

/db_xref="Subtilist:BG10502"

/db_xref="UniProt/Swiss-Prot:P37954"

/translation="MKDRFELVSKYQPGQGPAAIEKLVKIGQEKHQTLGATGTGKTFVSNLIKENVKPTLVIAHNKLAGQLSEFEFPNNAVEYFVSYYDYOPEAYVPQDTFIEKQASINDEIDKLRSATSLFERRDVIIIASVCIGYSGPEYEMVYSLRTEMERRELLRKVIDIOYANDIDFORGTNRVGDVVEIFPASRDEHCVVEFGDEIRIEVDALTGELIGDDHVAIFPASHFVTRAEKMEKAIQNIKELEBQKWNHNGKLEAQRLEQRTYDLEMRMGFCGSEKTSRHLLKPEGSTPTLLDYFPDDFMIVDESHVTIPQVGMFGNQARKQVVDGFRPLPSALDRPLRFEFEKHMNIYVSATGPYEIHTDEMVEQIIRPTGLDPLDVPIEGQIDDLIGEIQARIERNERVLVTTLRKMSEDLTVLKEIGIKVNIHSEITLERIEIIRDLRGKQVAGINLRGLIDIPESVAVIILADKSGPLRSBSLQITQGRARNAGEVIVYADTKTSMELAIERTKRBOOEERNHEHGTTPTKINKEIDYIRATVYADKAEYTKAPKLSKTKKERQKXVEQMEHMEKAKALDFERAELDLLELAEG"

/complement(5166..5396)

/gene="csbA"

/locus_tag="BSU35180"

/complement(5166..5396)

/gene="csbA"

/locus_tag="BSU35180"

/function="unknown"

/codon_start=1

/transl_table=11

/product="putative membrane protein"

/protein_id="CAB15535.1"

/db_xref="GI:2636044"

/db_xref="GOA:P37953"

/db_xref="Subtilist:BG10501"

/db_xref="UniProt/Swiss-Prot:P37953"

/translation="MITKAVPALFPPMLVPLFRVYFNNHYVALTRALLFASYLKGVTETVYFGLDVSLVAGLTMAKKAKEKEE"

/5843..8338

/gene="yvkC"

/locus_tag="BSU35190"

/5843..8338

/gene="yvkC"

/locus_tag="BSU35190"

/function="unknown"

/note="similar to pyruvate, water dikinase"

/codon_start=1

/transl_table=11

/protein_id="CAB15536.1"

/db_xref="GI:2636045"

/db_xref="GOA:O34796"

/db_xref="InterPro:IPR002192"

/db_xref="InterPro:IPR008279"

/db_xref="Subtilist:BG14114"

/db_xref="UniProt/T-EMBL:O34796"

/translation="MISVLRQAEBSQLAGAKAMNLIKTKHGLPYDPDFIOTNMLARFEMDQLOETSNVVEGGIISGTFSELDELGTSTSYKLRBESTRSVAVSSASBDLEGAFPAQCYEYFNLNKTBEERFLAKVCKMSFSGRGSYKKKKNQNIABPLMKIIVQGLISESGYVFSNPATPHDRLLIASYGLGAVGVATPPTFVINKSFEIOKEIGAKETYMESAAGSIAKETSEDMRSRCCLDQVIELAETTKTELDLYGVPVIERGIDAKQIYLLQARPTTIDQKKAEBKRSMTIDTDMNDMIMESNIIEGVSPLSFSTVPALEYGLEKSMQKPIGVVVDVVLVYGHYISKQGGQPPSEDCGKELPFIISLHMVDIINHNTYLPFYRTDOLAQETHEASLEAEOKLKATVLTAYEERPNVFPQILLTNKQAMVODIQGSENAHFYEMLTGRNKSLETRCLAMFSVEOVENPMLIAFENN

KPEOLKLEJEDTDEGRHFLKNVHEFLQEGMRSVSKHDLBQIWEVNPYPALANTONY
VRNGYHFNPEROKTEKREKEXKYNLEBSEIDBGLATREDEYXOWMTNSANTKDHFFY
IDAMLDARAFLEKIGELBENGVIOUREDIWLYDDEVEDALHAPLSLOKAEKR
QIFHEYLQAPAYAGTPTKQKAESEIVIDEKENTHIFGLAASSGIATGPV
KIRIDANESFOGAPGDLVCKMTPTPLMTSLFQDAKAIITDQGLISHAAILAREVGIP
AVLGRTRATERRDSDIITVDSSESKITVSRSS"
8414..8983

gene
/gene="yvkB"
/locus_tag="BSU35200"
8414..8983
/gene="yvkB"
/locus_tag="BSU35200"
/function="unknown"
/note="similar to transcriptional regulator (TetR/AcrR family)"
/codon_start=1
/transl_table=11
/protein_id="CAB15537.1"
/db_xref="GI:2636046"
/db_xref="GOA:O34643"
/db_xref="InterPro:IPR001647"
/db_xref="Subtilis:BG1413"
/db_xref="UniProt/TREMBL:O34643"

Query Match 98.8%; Score 442.2; DB 1; Length 200690;
Best Local Similarity 99.3%; Pred. No. 3.5e-104;
Matches 444; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGTCGATCAGATTATATACATCGACATATTTAGTGTACTCTAGTTAATTTT 60
DB 89478 ATGTCGATCAGATTATATACATCGACATATTTAGTGTACTCTAGTTAATTTT 89419
QY 61 GCGGAAAAACAGGATCGTCCGCGAGGACTTGTGTACCGGATATTTAGACTGTG 120
DB 89418 GCGGAAAAACAGGATCGTCCGCGAGGACTTGTGTACCGGATATTTAGACTGTG 89359
QY 121 TTTAATCAGCGGCTTTATTTACTGTTTCTAGTAGCTTGTCTACTATGTTATC 180
DB 89358 TTTAATCAGCGGCTTTATTTACTGTTTCTAGTAGCTTGTCTACTATGTTATC 89299
QY 181 GTGAATACGGTTATTCAGAAATTTATGATTTGTACGAGACGAGAAATTCGTCGCATG 240
DB 89298 GTGAATACGGTTATTCAGAAATTTATGATTTGTACGAGACGAGAAATTCGTCGCATG 89239
QY 241 CTGATACAGGAGATCGTCTTAAATCGCGTTGATTTTCTATACCGGATTTACCATTT 300
DB 89238 CTGATACAGGAGATCGTCTTAAATCGCGTTGATTTTCTATACCGGATTTACCATTT 89179
QY 301 GAAATCGCAGAAATTCGAGAAATCGGATCATGTCGACGTTTAATGCAATACCATTT 360
DB 89178 GAAATCGCAGAAATTCGAGAAATCGGATCATGTCGACGTTTAATGCAATACCATTT 89119
QY 361 CAGAAACAGGTTTAAACATTAAGTTCGAGAGACGCTGTATTTAGCGGAGGACCTTT 420
DB 89118 CAGAAACAGGTTTAAACATTAAGTTCGAGAGACGCTGTATTTAGCGGAGGACCTTT 89059
QY 421 GCTATCATGTTGTTTACTACTTAAT 447
DB 89058 GCTATCATGTTGTTTACTACTTAAT 89032

RESULT 6
E50426 450 bp DNA linear PAT 31-JAN-2002
LOCUS
DEFINITION Process for producing poly-gamma-glutamic acid.
ACCESSION E50426
VERSION E50426.1 GI:18629414
KEYWORDS JP 2001017182-A/3.
SOURCE
ORGANISM Bacillus subtilis
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE 1 (bases 1 to 450)
AUTHORS Ashiuchi, M., Misono, H. and Soda, K.

TITLE
JOURNAL
Process for producing poly-gamma-glutamic acid
Patent: JP 2001017182-A 3 23-JAN-2001;
NAGASE & CO LTD

COMMENT
OS Bacillus subtilis (hay bacillus) IFO 3336
PN JP 2001017182-A/3
PD 23-JAN-2001
PF 09-JUL-1999 JP 1999196335
PR

PI MAKORO ASHUCHI, HARUO MISONO, KENJI SODA
PC C12N15/09, C08G69/36, C12N1/15, C12N1/19, C12N1/21, C12N5/10, PC
C12N9/00, C12N9/90,
PC C12P13/14, C12N15/00, C12N5/00
CC
FT Key Location/Qualifiers
FT source 1..450
FT IFO 3336
/organism="Bacillus subtilis (hay bacillus)"

FEATURES
source Location/Qualifiers
1..450
/organism="Bacillus subtilis"
/mol_type="genomic DNA"
/db_xref="taxon:1423"

ORIGIN

Query Match 98.6%; Score 440.6; DB 6; Length 450;
Best Local Similarity 99.1%; Pred. No. 2e-103;
Matches 443; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATGTCGATCAGATTATATACATCGACATATTTAGTGTACTCTAGTTAATTTT 60
DB 1 ATGTCGATCAGATTATATACATCGACATATTTAGTGTACTCTAGTTAATTTT 60
QY 61 GCGGAAAAACAGGATCGTCCGCGAGGACTTGTGTACCGGATATTTAGACTGTG 120
DB 61 GCGGAAAAACAGGATCGTCCGCGAGGACTTGTGTACCGGATATTTAGACTGTG 120
QY 121 TTTAATCAGCGGCTTTATTTACTGTTTCTAGTAGCTTGTCTACTATGTTATC 180
DB 121 TTTAATCAGCGGCTTTATTTACTGTTTCTAGTAGCTTGTCTACTATGTTATC 180
QY 181 GTGAATACGGTTATTCAGAAATTTATGATTTGTACGAGACGAGAAATTCGTCGCATG 240
DB 181 GTGAATACGGTTATTCAGAAATTTATGATTTGTACGAGACGAGAAATTCGTCGCATG 240
QY 241 CTGATACAGGAGATCGTCTTAAATCGCGTTGATTTTCTATACCGGATTTACCATTT 300
DB 241 CTGATACAGGAGATCGTCTTAAATCGCGTTGATTTTCTATACCGGATTTACCATTT 300
QY 301 GAAATCGCAGAAATTCGAGAAATCGGATCATGTCGACGTTTAATGCAATACCATTT 360
DB 301 GAAATCGCAGAAATTCGAGAAATCGGATCATGTCGACGTTTAATGCAATACCATTT 360
QY 361 CAGAAACAGGTTTAAACATTAAGTTCGAGAGACGCTGTATTTAGCGGAGGACCTTT 420
DB 361 CAGAAACAGGTTTAAACATTAAGTTCGAGAGACGCTGTATTTAGCGGAGGACCTTT 420
QY 421 GCTATCATGTTGTTTACTACTTAAT 447
DB 421 GCTATCATGTTGTTTACTACTTAAT 447

RESULT 7
E50424 3045 bp DNA linear PAT 31-JAN-2002
LOCUS
DEFINITION Process for producing poly-gamma-glutamic acid.
ACCESSION E50424
VERSION E50424.1 GI:18629412
KEYWORDS JP 2001017182-A/1.
SOURCE
ORGANISM Bacillus subtilis
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE 1 (bases 1 to 3045)
AUTHORS Ashiuchi, M., Misono, H. and Soda, K.

TITLE Process for producing poly-gamma-glutamic acid
JOURNAL Patent: JP 2001017182-A 1 23-JAN-2001;
NAGASE ECO LTD
COMMENT OS Bacillus subtilis (hay bacillus) IPO 3336
PN JP 2001017182-A/1
PD 23-JAN-2001
PR 09-JUL-1999 JP 1999196335
PI MAKOTO ASHICHI, HARUO MISONO, KENJI SODA
PC C12N15/09, C08G69/36, C12N1/15, C12N1/19, C12N1/21, C12N5/10, PC
C12N9/00, C12N9/90,
PC C12P13/14, C12N15/00, C12N5/00
CC
FH Key location/Qualifiers
FT source 1..3047
FT /organism="Bacillus subtilis (hay bacillus)"
FT IPO 3336
FT Location/Qualifiers
1..3045
/organism="Bacillus subtilis"
/mol_type="genomic DNA"
/db_xref="taxon:1423"

ORIGIN
Query Match 98.6%; Score 440.6; DB 6; Length 3045;
Best Local Similarity 99.1%; Pred. No. 1.6e-103;
Matches 443; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 ATGTTGCGATCAGATTATATACATCGCACTAATTTAGGTGATCTACTGATTATTTT 60
DB 1215 ATGTTGCGATCAGATTATATACATCGCACTAATTTAGGTGATCTACTGATTATTTT 1274
OY 61 GCGGAAAAACAGGGATGTCGCGGAGGAGCTGTTGTAACGGGATTTTATAGACTTGG 120
DB 1275 GCGGAAAAACAGGGATGTCGCGGAGGAGCTGTTGTAACGGGATTTTATAGACTTGG 1334
OY 121 TTTATCAGCCGCTTTATTTTACTGTTTGTCTAGTAGCTTCTCACTATGTTATC 180
DB 1335 TTTATCAGCCGCTTTATTTTACTGTTTGTCTAGTAGCTTCTCACTATGTTATC 1394
OY 181 GTGAATACGGTTTATCCAAATTTATGATTTTGTACGAGCAGAAATTCGCTGCCATG 240
DB 1395 GTGAATACGGTTTATCCAAATTTATGATTTTGTACGAGCAGAAATTCGCTGCCATG 1454
OY 241 CTGATTAACAGGGATGTCCTTAATAATCGGCTTGAATTTCTATCCGATTTGATCAATTT 300
DB 1455 CTGATTAACAGGGATGTCCTTAATAATCGGCTTGAATTTCTATCCGATTTGATCAATTT 1514
OY 301 GAAATCGAGAAATTTGAGGAATCGGATCATCTGTCGAGGTTTATTTGCAATACAT 360
DB 1515 GAAATCGAGAAATTTGAGGAATCGGATCATCTGTCGAGGTTTATTTGCAATACAT 1574
OY 361 CAGAAACAGGTTTAAACATTAAGTTTGGAGACACGCTGCTATTTAGAGCGGAGCACTTT 420
DB 1575 CAGAAACAGGTTTAAACATTAAGTTTGGAGACACGCTGCTATTTAGAGCGGAGCACTTT 1634
OY 421 GCTATCATGTTGTTTACTACTTAAT 447
DB 1635 GCTATCATGTTGTTTACTACTTAAT 1661

RESULT 8
AB046355 3738 bp DNA linear BCT 11-JAN-2002
LOCUS AB046355
DEFINITION Bacillus subtilis ywsc, ywca, ywcb, ywcc genes, complete cds.
ACCESSION AB046355
VERSION AB046355.1 GI:13591556
KEYWORDS
SOURCE Bacillus subtilis
ORGANISM Bacillus subtilis
REFERENCE 1 Urushidate, Y., Tokuyama, S. and Tahara, Y.

TITLE Characterization of the Bacillus subtilis ywsc gene, involved in
JOURNAL gamma-polyglutamic acid production
J. Bacteriol. 184 (2), 337-343 (2002)
PUBMED 11751809
REFERENCE 2 (bases 1 to 3738)
AUTHORS Tahara, Y. and Urushidate, Y.
TITLE Direct Submission
JOURNAL Submitted (18-JUL-2000) Yasutaka Tahara, Shizuoka University,
Faculty of Agriculture, 836 Ohya, Shizuoka city, Shizuoka 422-8529,
Japan (E-mail: acyutah@agr.shizuoka.ac.jp,
Tel: 81-54-238-4878 (ex. 7808), Fax: 81-54-237-3028)

FEATURES
source
1..3738
Location/Qualifiers
/organism="Bacillus subtilis"
/mol_type="genomic DNA"
/db_xref="taxon:1423"
gene
CDS
497..1678
/gene="ywsc"
497..1678
/gene="ywsc"
/codon_start=1
/transl_table=1
/protein_id="BAB40948.1"
/db_xref="GI:13591557"
/translation="MMLLIACAVILVIGLEKRRQKIDALPRVNINGIRKSTV
RLTTLTGILAGYKVTGTTGDARMIMYDTPBEKPKRPOGPNIGQKXWRETV
RGANAIYSECAVAVPDYQIIFQELDQANIVYVNEIDMDVNGPLDIAEFTAT
IPVNGHVIDSEYTERFERKOKAKERNKVI IANSKITDSEYLRKFEVVPDNASL
GVADALIDETAFKMLNPPDGMARIIPLISPSRPGFVNGFANASSTLNTYK
RVEIGYPTDDPIITMNCRADVRDRIIPYELASLLIGQSEPIYKAYE
SKIPADKLHLEKYSTDEIMELKKRMHNVIVGVGINHGADELIERIHBYKQVLV
S"

gene
CDS
1693..2142
/gene="ywca"
1693..2142
/gene="ywca"
/codon_start=1
/transl_table=1
/protein_id="BAB40949.1"
/db_xref="GI:13591558"
/translation="MFGSDIYIALILGVLSLIPAEKTVIPAGIVYVYGLVFNOP
VFILVLVSLIVIVKYGSKFMILYGRKRPAMLTIGIVLKIADPLPIYFPI
AEFRGIIIVGLANTIOKGLITTFGSLLSGRAFAIMFYVYL"

gene
CDS
2161..3303
/gene="ywcb"
2161..3303
/gene="ywcb"
/codon_start=1
/transl_table=1
/protein_id="BAB40950.1"
/db_xref="GI:13591559"
/translation="MKKELSPHEKLKLTROQKKTNKHVFIPIPIVLFVAFPMAG
KAETPKVITSDVLSFVGDIMMGVYVEKTEQKADSIPOVEIIPFASDVAN
FENVTYQKNYKQADKEIHLQTNKESKVKLDNMFVLSANNHAMDYGVQKMDTIG
EPATNKLDIVAGAKSLDAKKKISYQVNGVLTATLGTFSVSGFAKNTPGVLA
DPBIFPMISEAKGADIIVVQSHMGSEYNDPDROROLARMSDGAIIIVGHNH
VLEPIEYVNGVTFYSLGNFVFDQGWRTSDALVOYHLKNGRGREVTPIDIHEAT
PAPYKDSLQKQTIIRLTKDSNPMKVBGKLTFLDIDHDKLSK"

gene
CDS
3318..3485
/gene="ywcc"
3318..3485
/gene="ywcc"
/codon_start=1
/transl_table=1
/protein_id="BAB40951.1"
/db_xref="GI:13591560"
/translation="MKFVKAIWPFVAVAVIFMFSAPKFNDDLTDEKOKIDMEWNTI
QQGEPEVYANAK"

ORIGIN
Query Match 97.5%; Score 435.8; DB 1; Length 3738;
Best Local Similarity 98.4%; Pred. No. 2.7e-102;
Matches 440; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

Qy      1  ATGTTGGATCAGATTATACATCCACTAATTTAGTGTACTGAGTTAATTTT 60
Db      1693 ATGTTGGATCAGATTATACATCCACTAATTTAGTGTACTGAGTTAATTTT 1752
Qy      61  GCGGAAAAACAGGGATGTCGCGGACGACTGTGTACCGGGATTTAGACTGTG 120
Db      1753 GCGGAAAAACAGGGATGTCGCGGACGACTGTGTACCGGGATTTAGACTGTG 1812
Qy      121 TTTATCAGCCGCTTTATTTACTTGTGTTGCTAGTAGCCTGCTCACTTATGTTATC 180
Db      1813 TTTATCAGCCGCTTTATTTACTTGTGTTGCTAGTAGCCTGCTCACTTATGTTATC 1872
Qy      181 GTGAATAACGTTTATCCAAATTTATGATTGTTGACGACGAGAAATTCCTGCGCATG 240
Db      1873 GTGAATAACGTTTATCCAAATTTATGATTGTTGACGACGAGAAATTCCTGCGCATG 1932
Qy      241 CTGATTAACAGGATGCTGCTTAATAATCGGCTTGATTTCTATACCCGATTTGACCATTT 300
Db      1933 CTGATTAACAGGATGCTGCTTAATAATCGGCTTGATTTCTATACCCGATTTGACCATTT 1992
Qy      301 GAAATCGCAGATTTGAGAAATCGGCATCATGTCGACGTTTAAATTCGCAATACCAT 360
Db      1993 GAAATCGCAGATTTGAGAAATCGGCATCATGTCGACGTTTAAATTCGCAATACCAT 2052
Qy      361 CAGAAACAAGGTTTAAACCATTAAGTTGGAAGCAGCTGCTATGAGCGGAGCGACCTTT 420
Db      2053 CAGAAACAAGGTTTAAACCATTAAGTTGGAAGCAGCTGCTATGAGCGGAGCGACCTTT 2112
Qy      421 GCTATCATGTTGTTTACTACTTAAAT 447
Db      2113 GCTATCATGTTGTTTACTACTTAAAT 2139

```

RESULT 9
AB017333_36/c
WPCOMMENT

Sequence split into 43 fragments LOCUS AB017333 Accession AB017333

Fragment Name	Begin	End
AB017333_00	1	110000
AB017333_01	100001	210000
AB017333_02	200001	310000
AB017333_03	300001	410000
AB017333_04	400001	510000
AB017333_05	500001	610000
AB017333_06	600001	710000
AB017333_07	700001	810000
AB017333_08	800001	910000
AB017333_09	900001	1010000
AB017333_10	1000001	1110000
AB017333_11	1100001	1210000
AB017333_12	1200001	1310000
AB017333_13	1300001	1410000
AB017333_14	1400001	1510000
AB017333_15	1500001	1610000
AB017333_16	1600001	1710000
AB017333_17	1700001	1810000
AB017333_18	1800001	1910000
AB017333_19	1900001	2010000
AB017333_20	2000001	2110000
AB017333_21	2100001	2210000
AB017333_22	2200001	2310000
AB017333_23	2300001	2410000
AB017333_24	2400001	2510000
AB017333_25	2500001	2610000
AB017333_26	2600001	2710000
AB017333_27	2700001	2810000
AB017333_28	2800001	2910000
AB017333_29	2900001	3010000
AB017333_30	3000001	3110000
AB017333_31	3100001	3210000
AB017333_32	3200001	3310000
AB017333_33	3300001	3410000

AB017333_34 3400001 3510000
 AB017333_35 3500001 3610000
 AB017333_36 3600001 3710000
 AB017333_37 3700001 3810000
 AB017333_38 3800001 3910000
 AB017333_39 3900001 4010000
 AB017333_40 4000001 4110000
 AB017333_41 4100001 4210000
 AB017333_42 4200001 422645

Continuation (37 of 43) of AB017333 from base 3600001 (AB017333 Bacillus licheniformis DC

Query Match 65.1%; Score 290.8; DB 1; Length 110000;
 Best Local Similarity 78.3%; Pred. No. 6.8e-65;
 Matches 349; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

```

Qy      1  ATGTTGGATCAGATTATACATCCACTAATTTAGTGTACTGAGTTAATTTT 60
Db      6638 ATGTTGGATCAGATTATACATCCACTAATTTAGTGTACTGAGTTAATTTT 66329
Qy      61  GCGGAAAAACAGGGATGTCGCGGACGACTGTGTACCGGGATTTAGACTGTG 120
Db      66328 GCGGAAAAACAGGGATGTCGCGGACGACTGTGTACCGGGATTTAGACTGTG 66269
Qy      121 TTTATCAGCCGCTTTATTTACTTGTGTTGCTAGTAGCCTGCTCACTTATGTTATC 180
Db      66268 TTTATCAGCCGCTTTATTTACTTGTGTTGCTAGTAGCCTGCTCACTTATGTTATC 66209
Qy      181 GTGAATAACGTTTATCCAAATTTATGATTGTTGACGACGAGAAATTCCTGCGCATG 240
Db      66208 GTGAATAACGTTTATCCAAATTTATGATTGTTGACGACGAGAAATTCCTGCGCATG 66149
Qy      241 CTGATTAACAGGATGCTGCTTAATAATCGGCTTGATTTCTATACCCGATTTGACCATTT 300
Db      66148 CTGATTAACAGGATGCTGCTTAATAATCGGCTTGATTTCTATACCCGATTTGACCATTT 66089
Qy      301 GAAATCGCAGATTTGAGAAATCGGCATCATGTCGACGTTTAAATTCGCAATACCAT 360
Db      66088 GAAATCGCAGATTTGAGAAATCGGCATCATGTCGACGTTTAAATTCGCAATACCAT 66029
Qy      361 CAGAAACAAGGTTTAAACCATTAAGTTGGAAGCAGCTGCTATGAGCGGAGCGACCTTT 420
Db      66028 CAGAAACAAGGTTTAAACCATTAAGTTGGAAGCAGCTGCTATGAGCGGAGCGACCTTT 65969
Qy      421 GCTATCATGTTGTTTACTACTTAAAT 446
Db      65968 GCTATCATGTTGTTTACTACTTAAAT 65943

```

RESULT 10
CP000002_36/c
WPCOMMENT

Sequence split into 43 fragments LOCUS CP000002 Accession CP000002

Fragment Name	Begin	End
CP000002_00	1	110000
CP000002_01	100001	210000
CP000002_02	200001	310000
CP000002_03	300001	410000
CP000002_04	400001	510000
CP000002_05	500001	610000
CP000002_06	600001	710000
CP000002_07	700001	810000
CP000002_08	800001	910000
CP000002_09	900001	1010000
CP000002_10	1000001	1110000
CP000002_11	1100001	1210000
CP000002_12	1200001	1310000
CP000002_13	1300001	1410000
CP000002_14	1400001	1510000
CP000002_15	1500001	1610000
CP000002_16	1600001	1710000
CP000002_17	1700001	1810000
CP000002_18	1800001	1910000
CP000002_19	1900001	2010000

CP000002_20 2000001 2110000
CP000002_21 2100001 2210000
CP000002_22 2200001 2310000
CP000002_23 2300001 2410000
CP000002_24 2400001 2510000
CP000002_25 2500001 2610000
CP000002_26 2600001 2710000
CP000002_27 2700001 2810000
CP000002_28 2800001 2910000
CP000002_29 2900001 3010000
CP000002_30 3000001 3110000
CP000002_31 3100001 3210000
CP000002_32 3200001 3310000
CP000002_33 3300001 3410000
CP000002_34 3400001 3510000
CP000002_35 3500001 3610000
CP000002_36 3600001 3710000
CP000002_37 3700001 3810000
CP000002_38 3800001 3910000
CP000002_39 3900001 4010000
CP000002_40 4000001 4110000
CP000002_41 4100001 4210000
CP000002_42 4200001 4222334
Continuation (37 of 43) of CP000002 from base 3600001 (CP000002 Bacillus licheniformis A

Query Match 65.1%; Score 290.8; DB 1; Length 110000;
Best Local Similarity 78.3%; Pred. No. 6.8e-65;
Matches 349; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

Qy 1 ATGTTGGATCAGATTATACATCGACCTAATTTAGTGTAAGTCTAGTTAAATTTT 60
Db 66228 ATGTTGGATCAGATTATATATCCCTCATTTTAGAGTCTTACTCAGTTGATTTT 66169
Qy 61 GCGGAAAAACAGGATCGTCCGCGCAGAGACTTTGTATACCGGGAATTTTAGAATTGTG 120
Db 66168 GCAGGAAAAACGGAAATTGTACACCGCCCTCGTACCGGTTATTTGGAGCTTGTC 66109
Qy 121 TTATATCAGCCGGCTTTATTTTACTTTTGTAGTAGCTTCTCATTATGTTATC 180
Db 66108 TTCAATCAGCCGATTTTCAATGCTGCTCTTTTGTATGATTTCTAGCTATGTCATC 66049
Qy 181 GTGAATACGGTTTATCCAAATTTATGATTTTGTACGAGCAGAAAAATTCGCTGCCATG 240
Db 66048 GTGAATTCGGACTTTCAGAAATTTATGATTTGTACGAGCAGAAAAATTCGCAATG 65989
Qy 241 CTGATTAACAGGATCGTCTTAAATAATCGCGTTTGTATTTCTATACCCGATTTGACATT 300
Db 65988 CTGATTAACGGGAATCTTTTGAATTCGTTTGTATTTATATCCGATGCGCTTT 65929
Qy 301 GAAATCGAGAAATTCGAGAAATCGGATCATCGCCAGGTTTAAATGCGCAATTCATT 360
Db 65928 GAGATTCGCGAATTCAGGGGAATCGGAATCATCGTCCGGGCTATGCGCAATTCATT 65869
Qy 361 CAGAAACAGGTTTAAACATTACGTTCCGAGACAGCTGCTATTTAGCGGAGCACTTT 420
Db 65868 CAAAGACAGGATTAACGATTAACGCTTGAAGTACGCTTTATTTAGCGGAGCAACATTC 65809
Qy 421 GCTATCATGTTTGTACTACTTAAT 446
Db 65808 GTCAATTATGATGCTTACTACTTAAT 65783

RESULT 11
LOCUS AX34510 450 bp DNA linear PAT 28-JUN-2002
DEFINITION Sequence 2925 from Patent WO0229113.
ACCESSION AX34510
VERSION AX34510.1 GI:21659318
KEYWORDS
SOURCE Bacillus licheniformis
ORGANISM Bacillus licheniformis
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE 1

AUTHORS Berka, R. and Clausen, I. G.
TITLE Methods for monitoring multiple gene expression
JOURNAL Patent: WO 0229113-A 2925 11-APR-2002;
Novozymes Biotech, Inc. (US) ; Novozymes A/S (DK)
FEATURES
source Location/Qualifiers
1..450
/organism="Bacillus licheniformis"
/mol_type="unassigned DNA"
/db_xref="taxon:1402"
ORIGIN
Query Match 64.7%; Score 289.2; DB 6; Length 450;
Best Local Similarity 78.0%; Pred. No. 3.6e-64;
Matches 348; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

Qy 1 ATGTTGGATCAGATTATATCATCGACTAATTTAGTGTAGTACTCAGTTAAATTTT 60
Db 1 ATGTTGGATCAGATTATATATCGCTCATTTTATAGAGTCTTACTCAGTTGATTTT 60
Qy 61 GCGGAAAAACAGGATCGTCCGCGCAGAGACTTTGTATACCGGGAATTTAGACTTGTG 120
Db 61 GCAGGAAAAACGGAAATTGTACACCGGCTCGTACCGGTTATTTGGAGCTTGTC 120
Qy 121 TTATATCAGCCGGCTTTATTTTACTTTTGTCTAGTAGCTTGTCTACTTATGATC 180
Db 121 TTCAATCAGCCGATTTTCAATGCTGCTGCTTTTGTCTAGTTGCTGAGTATGATC 180
Qy 181 GTGAATACGGTTTATCCAAATTTATGATTTGTAGAGCAGCAAAATTCGTCGCATG 240
Db 181 GTGAATTCGGACTTTCAGAAATTTATGATTTATACGAGCAGCAAAATTCGCAATG 240
Qy 241 CTGATTAACAGGATCGTCTTAAATAATCGCTTGAATTTTCTATACCCGATTTGACATT 300
Db 241 CTGATTAACGGAATCTTTTGAATAATCGTTTGTATTTATATACCGGATGCCGTT 300
Qy 301 GAAATCGAGAAATTCAGAGATTCGATCATCTGCGCAGGTTTAATGCCAATACATT 360
Db 301 GAGATTCGCGAATTCAGGGGAATCGGAATCATCTGCGGCGTGAATCGCAATACATT 360
Qy 361 CAGAAACAGGTTTAAACATTACGTTGGAAGCAGCGTGCATTTAGCGGAGCACTTT 420
Db 361 CAAAGACAGGATTAACGATTAACGCTTGAAGTACGCTTTTATTTAGCGGAGCAACATTC 420
Qy 421 GCTATCATGTTTGTACTACTTAAT 446
Db 421 GTCAATTATGATGCTTACTACTTAAT 446

RESULT 12
LOCUS BACCAPABC 3244 bp DNA linear BCT 26-APR-1993
DEFINITION B.anthraxis encapsulation protein genes (capA, capB, and capC),
complete cds.
ACCESSION M24150
VERSION M24150.1 GI:142630
KEYWORDS encapsulation protein; membrane-associated protein.
SOURCE Bacillus anthracis
ORGANISM Bacillus anthracis
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus; Bacillus
cereus group.
REFERENCE 1 (bases 1 to 3244)
AUTHORS Makino, S., Uchida, I., Terakado, N., Sasakawa, C. and Yoshikawa, M.
TITLE Molecular characterization and protein analysis of the cap region,
which is essential for encapsulation in Bacillus anthracis
JOURNAL J. Bacteriol. 171 (2), 722-730 (1989)
COMMENT
PUBMED 2536679
ORIGINAL source text: B.anthraxis (strain TE702; isolate pCAP1)
DNA.
Draft entry and computer-readable sequence for [1] kindly provided
by I.Uchida, 01-MAY-1999.
FEATURES
source Location/Qualifiers
1..3244
/organism="Bacillus anthracis"

gene
CDS
/protein_id="AA26164.1"
/db_xref="GI:20520283"
/translation="MKKKRYKXHPRIQSAHYHIIQNAVYLVNAVAAHTI"
complement (1746..2594)
/gene="BXB0004"
complement (1746..2594)
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein, (px02-05)"
/protein_id="AA26165.1"
/db_xref="GI:20520284"
/translation="MSIKFNNKGRKPAKTRKAYLPTLGSFVTFPVLCTSPFSG
ATYPEEMKIGEVOSISTVKIAVAKKRNPDNOLRIYDELADNDOSILSNKKYK
ENKTIKQKDNVKTQYRASPNYIVISENPPEEGVSVYKPEYIHPQLQNVDDL
KERMKVYLENEKLIINELKKSDFERYELAFSQALKEIEKKEDSSAMKOL
KIKNEOLTKEMEYTGEEKVKTQINSNESTINNHQKEIDLKEDIKMEKKIQLLD
EKKTI"
complement (2591..2866)
/gene="BXB0005"
complement (2591..2866)
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein, (px02-06)"
/protein_id="AA26166.1"
/db_xref="GI:20520285"
/translation="MLISELPHNDYAKGKYLILSLRKKKGVKDKFIEPIPTFDY
NFGKEVWLIVREYQPSVNGKRYNCLENIHAQSVVNVGSRMK"
complement (2955..3563)
/gene="BXB0006"
complement (2955..3563)
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein, (px02-07)"
/protein_id="AA26167.1"
/db_xref="GI:20520286"
/translation="MNNLSKILSGALVLSFNGIWAYNTTKEPKDKKXRVTVSEK
EKQIGDLVKLEQKIKINEGAGENKKEGSTLDLQKREYVANOQFHAVIDSVON
KGERNNLKITDKKVIDIVAPNTDLDLPNPKSVNKAIIYINSEGVSKKCTALLD
IEYIEGLENKQTTINSYVKITLEKQSEIVVENPYVPR"
complement (3589..4731)
/gene="BXB0007"
complement (3589..4731)
/note="identified by match to PFAM protein family HM
PF00877"
/codon_start=1
/transl_table=11
/product="NLP/P60 family protein, (px02-08)"
/protein_id="AA26168.1"
/db_xref="GI:20520287"
/translation="MAAORTAIDYKKYKRIKIVRLGGSTGVVIAAATILLIIVS
MAIFGGQSTGPNQGISGATVKNLPPEVNRQMAVQECAGQVEPELVYLAIIIM
VESNGISEKLPIQSSSQGMANTTINPADSIYGVMLKGAFDKMLGINDLLA
IVQYINFGNRYVHMLAANKKTHSIOTADYSLIVVAPAGNRNGTTIGSQPVAVAN
GGYRIRNGNFFYAEVEMKQVSLPDAGTSGQIPGSETPKMMDEVLYKNGNPYVWG
KSSSGQSGGLTYWAKTAGITIPISAALOYDPTVADKDAQPGDLVFPRTYGG
PNHSHVGIYDANTMDSNGSVGYHPTSSYMQHAYGIRRYR"
complement (4731..6667)
/gene="BXB0008"
complement (4731..6667)
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein, (px02-09)"
/protein_id="AA26169.1"
/db_xref="GI:20520288"

gene
CDS
/translation="MWLGGKEQKRTGEGYDLDIFSRVQPGGKIKENRYIAQDCY
VACILVYSIAEDIEPLMTALMNNKDTISSVDATANKREVVDINSITELXDRMTS
ESCTDRDDAHMELONLTPARSITQGGSEIVLVKRIYIDPLVLEGLRIGDIKKE
IAGNOAYVYVVKQKEKQTLFASVDOIEYGVSGYPLPSKNIGFGIYPHNDOK
DPREYVIGQSTGCAPIILDPFSTGRITSSFGKMGAKSTLKLQREGIYAKDC
FIRGEDKARDYTVVQGGKIIDLSGVQSEEVETGMINPLEVFATKITNTGAVDK
GSFIOHISKVTNMIPLNTEFDVTVIDEFKRLYAFETEMGLVPQKSDPYKVTGP
PNHYPILEDYKYSNLKEGATPQHRLELAIKQVEDMITVYGMFNGHTLKNP
ENEOIYLEDIDIGISKDYVNCOLFALTLIINSHALNGROMKYLEEKSLIEDVK
YFMVLEDCNIVINSONEFANRYVLEEREMRKFPAGI PFATOSPOEMLPPGSTVY
STMAKATIELQYKIFLMDNSVLETLKSUVGESITSEBRILPELRGELAVQVSTTE
TYNVFDPDEKQSERFGQ"
complement (6684..7358)
/gene="BXB0009"
complement (6684..7358)
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein, (px02-10)"
/protein_id="AA26170.1"
/db_xref="GI:20520289"
/translation="MKKRVATIAEAKKVRKROISEKSNQKOFISTIFASFKTQDAS
EIVDIAPYKRLQKGTGLVDRNNLQYLVKKTDTLSDNDLKRFNOLITSLRVH
EPFKILSTVSTERTBOOVYWKRALRFQGRMSQVSEKKEHLMYRYSALLENLR
VLWVEKMLKELAFPIVYVGNKTELIIKNVKOMKRYGGRQPNLQNMKAKEVEKLI
NMNSEW"
complement (7416..7730)
/gene="BXB0010"
complement (7416..7730)
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein, (px02-11)"
/protein_id="AA26171.1"
/db_xref="GI:20520290"
/translation="MEMNRNPKNTKOEIKLAFYIIDGAIILMLVLASVMPKVPVY
GFGRIYFVYLGFTGLFLCIKPHNSPFRNIFVILDWLKMNDKNYHPIEVNTISSEK
RK"
complement (7733..7936)
/gene="BXB0011"
complement (7733..7936)
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein, (px02-12)"
/protein_id="AA26172.1"
/db_xref="GI:20520291"
/translation="MNNWIFAILMLGVAIVLSIATFPKQIFTFKRRKKKEPVWTV
TNAEYLDKDPFKKQKQKKEHD"
complement (7938..8192)
/gene="BXB0012"
complement (7938..8192)
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein, (px02-13)"
/protein_id="AA26173.1"

Query Match 53.1%; Score 237.4; DB 1; Length 94829;
Best Local Similarity 70.7%; Pred. No. 4.8e-51;
Matches 316; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 1 ATGTTGGATCAGATTATACATGCACTAATTTAGGTGATCTACTACGATTATATTTT 60
|||||
DB 55584 ATGTTGGATCAGATTATATATATGATTAAGAGTTACAGTACGAGCTATTTT 55525
|||||
QY 61 GCGGAAAAACAGGATGTCGCCGACAGACTTGTGATCCGGATVATTGACACTGTG 120
|||||
DB 55524 ACAGAAAGAACAGGATTTTACTGACAGGTTAGTGTGACCTGTTATTGACACTCGTT 55465
|||||

OY	121	TTTAAATCAGCGGCTTTATTACTGTGTTTCCTGTGAGCGTGCTCACTATGTTATC	180
Dd	55464	TTTTAATCAGCCCGGATTTATGTTGGTTGTTTATTATCAGTATTTTAACATATGTAATC	55405
OY	. 181	GTGAATAACCGGTTTATCCAAATTTATGATTTTGTACGAGCGAGAATAATTCGCTGCATG	240
Dd	55404	GTTAACGATAGTGTGTTTCAAATTCATGATTTTATDAGCGCGTAGAAAATTTGGCGCAAG	55345
OY	241	CTGTATAACAGGAGATCGTCTTAAAAATCGCGTTTGATTTTCTATACCAGATGACATTT	300
Dd	55344	CTAATTACAGGATATTGTTTAAAACTTTTATGATTTATGTTATCTGTATGACCATTT	55285
OY	301	GAATGCCAATAATTGAGGAATCGGCATATCGTGCCAGGTTTAATTTGCCAAATACATT	360
Dd	55284	GAGATTTTGAATTCCTGTGATTTGAGATTATGTTCCAGAGATTAATTTGCAATACATT	55225
OY	361	CAGAAACAAGGTTTAAACCATTAACGTTTGGAGAAGCACGCTGATATGACGGAGGACCTTT	420
Dd	55224	CAAGAACAAGGTTTACCATTAACATTAACATTTGAAGTAACTTGTAAAGTGTGCAACATTT	55165
OY	421	GCATCATGTTTGTGTTTACTTAAATT	447
Dd	55164	GCATCATGAATATTATTACTTAATTT	55138
<hr/>			
RESULT_14	AE017335/c	94830 bp	DNA circular BCT 09-JUL-2004
LOCUS	AE017335	Bacillus anthracis str. 'Ames Ancestor'	plasmid pXO2, complete sequence.
DEFINITION	AE017335	AE017335	GI:50118566
ACCESSION	AE017335	AE017335	GI:50118566
VERSION	AE017335.3	GI:50118566	
KEYWORDS			
SOURCE			
ORGANISM			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REMARK			
COMMENT			
FEATURES			
source			
gene			
CDS			

gene
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein, (px02-03)"
/protein_id="AAI28931.2"
/db_xref="GI:47552343"
/translation="MTYVYLFQFRGNQIHVDIVQDFQRYPEYREADVLRHAIYLLD
EKRRBGGQDVLSLKKKEVDLQSYVNTQCOVDYLIKNMLIAQSGSLSEKLEQ
EVNSIASVYTKSEGGKFPRTVKREPIEENVQSHQIQREYKTSYVKKPLPIITDVG
QIVKKKGPPIQVYFRLGTYQGEIEMEQIPEHRKELE"
complement (1151..1438)
/locus_tag="GBAA_px02_0002"
/locus_tag="GBAA_px02_0002"
complement (1151..1438)
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein, (px02-04)"
/protein_id="AAI28932.2"
/db_xref="GI:47552344"
/translation="MEIHRNVDPDPIHLKIDKCKEIKGLKGRYRWEYINMMFEQH
FNQYSNKKDEKPEAVSNVSTIDLRQSDKLOEYDATHLVAAMIKLNE"
complement (1652..1768)
/locus_tag="GBAA_px02_0003"
/locus_tag="GBAA_px02_0003"
complement (1652..1768)
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAI28933.2"
/db_xref="GI:47552345"
/translation="MKKKRKRYKVHPRRISATRHQINAYELKNAYKATI"
complement (1746..2594)
/locus_tag="GBAA_px02_0004"
/locus_tag="GBAA_px02_0004"
complement (1746..2594)
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein, (px02-05)"
/protein_id="AAI28934.2"
/db_xref="GI:47552346"
/translation="WSIKFPMNKKGRKPANTKRAIYPTLGSFYTFMFVLCISVPSGS
ATYRKKEKIGEYQSLSTVYKIAVAKEVNDQNTLIDELRADNSQISNNKRYK
ENKYKQCNQNDKTKYVRASDNYIVVISENVPEFGVSVVKEVYIHPQLQNDVDL
KERSKMYVLENEKLINRELKCKSKDFERYELAFSOALRKEIEKMEBKSAMKQID
KIKNQLKMEVYQTEGKVKTKNTINSNESTINNHQEIDVLKEDIKMEKKIQLLD
EKRTI"
complement (2591..2866)
/locus_tag="GBAA_px02_0005"
/locus_tag="GBAA_px02_0005"
complement (2591..2866)
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein, (px02-06)"
/protein_id="AAI28935.2"
/db_xref="GI:47552347"
/translation="MLSELKPNHDVAKBGKYLILSRKKKGVKDKFIEPIITWFDV
NFGREVYMLVREYQPSYNGKEXYTNCKLENIHQAQSVVNVKGRMK"
complement (2955..3563)
/locus_tag="GBAA_px02_0006"
/locus_tag="GBAA_px02_0006"
complement (2955..3563)
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein, (px02-07)"
/protein_id="AAI28936.2"
/db_xref="GI:47552348"
/translation="MNNLSKILSGALVLSFNGIWAYNTTKEFDKDKRYRVVSEK
ENQIGDLVKYLEQDKKINNEGAGENKKEGNSTLIDLQKKRYEVANQFVHAYLVDVSYQ

[illegible]

Query Match	Best Local Similarity	70.7%	Score 237.4	DB 1	Length 94830
Matches 316	Conservative 0	Mismatches 131	Indels 0	Gaps 0	
QY	1	ATGTTCGGATCAGATTATATACATGACATATTTAGTGTACTACTGCTTAATTTT	60		
Db	55584	ATGTTCGGATCAGATTATATATATATGATTAAGTATGAGATTAACCTGAGCCCTTAATTTT	55522		
QY	61	CGCGAAAAAACAAGGATTCGTCGCGGAGAGCTGTGTACCGGATATTTAGACTGTG	120		
Db	55524	ACGAAAAACAAGGATTTTCTACCTGAGGTTTATGTTGACCTGTTATTTAGCACTCGTT	55465		
QY	121	TTTAATCAGCCGCTCTTATTTTACTTGTGTTGCTAGTACGCTTCTATATGTTATC	180		
Db	55464	TTTAATCAGCCGCTATTTATGTTGTTTATTTATCAGATTTTAAACATATGTTATC	55405		
QY	181	GTGAATACGCTTTATTCATTAATGATTTGTATGACGACGACGAAAAATTCCTGCCATG	240		
Db	55404	GTTACGATATGTTTCAAGATTCATGATTTTATGATGACCGGATGAAAAATTCGCGCAACG	55345		
QY	241	CTGATTAACAGGATTCGCTCTTAAAAATCCGCTTGTATTTCTATACCCGATTTGACATTT	300		
Db	55344	CTAATTAACAGATTTGTTTAAAACTTTTATTTGATTTATGTTATTCCTGTATGACATTT	55285		
QY	301	GAATCGCAGAAATTTTGAGAGAAATCGGCATCATCGTCCAGAGTTTATTCACATACATT	360		
Db	55284	GAGATTTTGAATTCCTGCTGATATGAGATTTATGTTCCAGATTAATTCGAAAAATCAATT	55225		
QY	361	CAGAAACAGGTTTAAACATTAATCGTTGGAAGACGCTCTATTTGAGCGAGGACCTTT	420		
Db	55224	CAAAAGCAAGGTTTACCATTAACATTTGGAACATTAATTTGTTAAGTGTGCAACATTT	55165		
QY	421	GCTATCATGTTTGTATCTACTTAAT	447		
Db	55164	GCAATCATGATTAATTAATTAATTAATTT	55138		
RESULT 15	AF188935/c				
LOCUS	AF188935	96231 bp	DNA	circular BCT 01-OCT-2003	
DEFINITION	Bacillus anthracis plasmid pX02, complete sequence.				
ACCESSION	AF188935				
VERSION	AF188935.1	GI:6470151			
KEYWORDS					
SOURCE					
ORGANISM	Bacillus anthracis				
	Bacillus anthracis				
	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus; Bacillus cereus group.				
REFERENCE	1. (bases 1 to 96231)				
AUTHORS	Ohtaka, R.T., Cloud, K., Hampton, O., Hill, K.K., Keim, P., Lamke, G., Kumano, S., Mancer, D., Martinez, Y., Svensson, R., Tatum, L.R., Brown, A.E. and Jackson, P.J.				
TITLE	Direct Submision				
JOURNAL	Submitted (01-NOV-1999) Bioscience Division, Los Alamos National Laboratory, M888, Los Alamos, NM 87545, USA				
FEATURES					
source	1. 96231				
	/organism="Bacillus anthracis"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:1392"				
	/plasmid="pX02"				
	complement(19..477)				
	/note="hypothetical protein"				

/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PX02-01"
/protein_id="AAFI3606.1"
/db_xref="GI:6470152"
/translation="MRGKGGDDVSLKKEVKDKOSVNTMKQMDVLIKIMELIAO
SGSNLSLEKLEQVKSLEIASSVTKSEKFPRTVYKREIEENVQEHQIQREVKTTS
VKPLPMITDVQIVMKKGKPHQVPTRLGTGYQIEWEQIPEHRRKELE"
complement(482..1018)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PX02-02"
/protein_id="AAFI3607.1"
/db_xref="GI:6470153"
/translation="MGKSPSITLTSQYITLPPNKSFTYVYDQCMROKALLKDGHTLNT
BITLSRLQNAIDSPDMERGSPHSQKYSVEKAEKAVLLTSNSFTTDDPKYV
SNWGRKVALSKKKTKEEBEVSLLNKHITDLSPEMKEGKKEGIKAEIISGVFSK
DKNVITSTDNLNKKITIR"
complement(1037..1633)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PX02-03"
/protein_id="AAFI3608.1"
/db_xref="GI:6470154"
/translation="MTYKLFQFRGNQIHVDIVDFORQYPPREADVLRHATVLLD
EKRRSGQDDVSLKKEVKDKOSVNTMKQMDVLIKIMELIAOSSGNLSLEKLEO
QVWKSSVTKSEKFPRTVYKREIEENVQEHQIQREVKTTSVKKPLMITDVQ
IVMKKGKPHQVPTRLGTGYQIEWEQIPEHRRKELE"
complement(1645..1932)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PX02-04"
/protein_id="AAFI3609.1"
/db_xref="GI:6470155"
/translation="MEIHRYNDPPLYLKEIDRCKEIGKKIGRRYRMEYIMMPEQH
FQOYSRNKEDFDEAVSNVSTTLDRQSKLOEYIDATIELVAAMTKLNEE"
complement(2240..3088)
/note="similar to Dicyostelium discoideum
glutamine-asparagine rich protein encoded by Genbank
Accession Number U07817"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PX02-05"
/protein_id="AAFI3610.1"
/db_xref="GI:6470156"
/translation="MSIKFRMNKGGKRPANTRKAYILFTLGSPTMFEVLCISPVESG
ATYKPEMKIGEQSLSTYKLAIVAKKEINPDQTLRIYELPADNDQSILSMKRY
ENKYIKQNDNVKTVKVRASDNYIVISENPEEFVSSVAVPEYIHPQLQNDVDL
KERSMKVYLVENEKLIINRELKKSQFYEREYIAFQOALRKEIEKMEKSSAMQOL
KIKNEQLTKEMEYQTEGEKVTKNITINSNESTINNHOKEIDVLKEDIKMEKKIKQLD
EKKKT"
complement(3085..3360)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PX02-06"
/protein_id="AAFI3611.1"
/db_xref="GI:6470157"
/translation="MLSELKPNHDYAKESGKYLILSRKKKGVKDKFTLEIPITWFDY
NPEKRYEWLIIVREYQPSVNGEKYTNCKLENINHAQVSVVAVKGERMK"
complement(3449..4057)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental

/transl_table=11
/product="PX02-07"
/protein_id="AAFI3612.1"
/db_xref="GI:6470158"
/translation="MNNLSKILSGALVLSFNGCIWAVNTTKEPDKDKRYRVYSEK
EKOIGDLKVLQEDKKINEGAGENKEGSTLIDLNQKREVAHQFPAHLDYSVN
KEGRNNILKTTDKKVDIVAPNTDDIDDPFKSHVKKAAIYINSEGVSKCTYALD
IEYITGLENKQTTINSVKTILEKQEEIKVEYNYPYVR"
complement(4083..5225)
/note="similar to Enterococcus faecalis transposon TN916
ORF14 encoded by Genbank Accession Number AB001488"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PX02-08"
/protein_id="AAFI3613.1"
/db_xref="GI:6470159"
/translation="MAATQETAIDKKYKRIKMIIVLLGGSTGVVIAAATLLIIVS
MAIRGGQSTGTPNGGISGATVKNLPEVNRWQAMQEQEAAQGVPELVYVLAIM
VESNGISEKLPIDIQSSQSGAMANTISNPQSIYGVMLHKAQFADAKMIGINDLA
IVQYVNRGRVHMLANNKTHSIOTADYSLTVAPAGRNRTTIGYQPVAVAN
GGIRYINGANFPYAMVQYLSFDGAGTSGQISGSETFVYMDDEVLYKNGNPVYWG
GKSSQGFDCSGLTVMAYKTIGITIPISAATQPTVAVDPKDAQPGLVFPRTYGG
PNHVSIVGIYIDANTMYDSNGSGVGHQFTSSVQOHVAGIRRVPR"
complement(5227..7158)
/note="similar to Staphylococcus aureus transfer complex
protein trsb encoded by Genbank Accession Number L11988"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PX02-09"
/protein_id="AAFI3614.1"
/db_xref="GI:6470160"
/translation="MKLSKGGQKRTCTGYLDPTISRQPGGIKFNENKIYAQGCYV
ACIHVYSLAEDIPLMLTALNNKDTISSVVAATNKEVKVDIKRSITELKDRPTE
RSTGRDADHMLQNLTDPARSITQGGIIVKLVKRIYIPVLYQLKRIQDIIKEI
AGONKAQVAVKQKEWQTLFASDQIEYLVKSGVPLSKNIGFQIPFHQDLKD
PRGYLGQTSQGAFTLDPSPSTGRTSFSGIPGMAAGSTLTKOLEGVAADCF
IRGPKARDYITVVOQSGKITLDSGVQDSEVEVGMNPLPEVATKLTNNGAVDEK
SFIQHSIVTNNIRPLNTEPDVTYIOERKHLIYATITKGLVPQKGSRPYVYVYCP
NHYPILBDFYKYLINLKLGAQTPQHRHDLKIQVDMITIVYDMEFNGLTTLKFE
NEQIVFPIIDISKYDKSVFNCQLTALTALIMSHALKNGRQMKYLRREKNLSIEVYK
TMKPLDECHVYINSQNEFANRVLYEFSEMRKFPAGIPEATQSPQEMLPSPSTYDVS
TMKAPILFQVYKIPLMNDSVLETLSVLSGSLTSEFRILPELRGEALIVQVSTET
XNMFDPDEKQERKGGQ"
complement(7178..7846)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PX02-10"
/protein_id="AAFI3615.1"
/db_xref="GI:6470161"
/translation="MKVATAEAKVNRKQKQISEKSNQSGSTSTIFASFNKTASIV
DIAPYKRLQKQIGLVDHNNLOVYLKVYTTDLLSNQODDLKRFPMQLSLCRVYHEP
PKILSTYETETEOOVVKKPMALRFQSGQSEVSEKKEHLMYQRYSLAINLRVL
VWEKNIKELAFITVYKQNETLILNKVADMKRYGGRQPLQMKAKYEVKILFKLOM
NSEM"
complement(7910..8224)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PX02-11"
/protein_id="AAFI3616.1"
/db_xref="GI:6470162"
/translation="MEMMENPKNTQOEIKLAFYIIDGAILALMEVLASYMPKVPVG
GGRIMFYLVFGTFGLFLCIKRNSTPFRNNTVIIDMLKMDKKNHPIEIVNTISSETK
RK"
complement(8227..8430)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental

CDS

CDS

```
/transl_table=1
/product="PX02-12"
/protein_id="AAFI3617.1"
/db_xref="GI:6470163"
/translation="MNNWIFAILMLGVAIVLSIATPFKQIFTIKRRKKPEVWTV
TNAEYLDIDFWKQKQDKKEHD"
complement(8432..8686)
/feature="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="PX02-13"
/protein_id="AAFI3618.1"
/db_xref="GI:6470164"
/translation="MKVIDIARRRIYFPMQKQELRASILMTVAGFIIFAILVFOIS
FELGHLVHYIVFAFLYLSLHLYSNNKLARKIEKQGY"
complement(8704..11562)
/feature="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="PX02-14"
/protein_id="AAFI3619.1"
/db_xref="GI:6470165"
/translation="MATMKIKEMFPOSKKALVIGILAVITLFVSFTFPPIAVAFDWE
FLGKVDTEKRVAFLOEMIOYADFLGISWISOGGIIKTLIVLSALEGLIDPTFS
FFDIKDGANDPASIIKGLFYALFVVIWMLGIRTIQHKRPRKSVGNILMIG
LLGGLNELADMDQKSTIDFSEITITSKADGLANDLVKONTADILYISKTFEPIOS
KKNEPVAISSTDSKRVKNSKDIPLKQGLGVVPKVIEMLSQDIPKETEYLVYKI
TNDGKETVEKINDSVFNPFDKFPFGYVRYPMNGTIFMGLALGVAVLFVVFVFM
```

Query Match 53.1%; Score 237.4; DB 1; Length 96231;

Best Local Similarity 70.7%; Pred. No. 4.8e-51; Matches 316; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

```
QY 1 ATGTCGATGAGATTATACATCGACTAATTTAGGTACTACTAGTTAATTTT 60
DB 56074 ATGTTGATCGATTATATATGCAATTAGAGTTAGACTGAGCTTATTTT 56015
QY 61 GCGGAAAAACAGGATCGTCCGCGAGACTGTTGTACCGGATATTAGACTGTG 120
DB 56014 ACAGAAAGAACAGATTATCTGAGGTAGTTGATCTGTTATTAGACTCGT 55955
QY 121 TTTATCAGCGGCTTATTTACTGTTTGTAGAGCTGCTCACTTATGTATC 180
DB 55954 TTTATCAGCCCGATTTATGTTGTTGTTTATTATCAGATTATTACATATGATC 55895
QY 181 GTGAATACGTTTATCCAAATTTATGATTTTGTACGACGAGAAAATTCGTCGATG 240
DB 55894 GTTACGTATGTTGTTCAAGATTCAATTTATATGCGCGAAGAAATTTGCGGCAAG 55835
QY 241 CTGATTAACAGGATCGTCCAAAATTCGCGTTGATTTTCTATACCGGATTTACATTT 300
DB 55834 CTAAATTACAGGATTTGTTTAAACTTTTATTTGATTTGTTATCTGTTATGCCATTT 55775
QY 301 GAAATCGAGATTTCAGGAATCGGATCATCGTCCAGTTTAAATGCGAATACCAT 360
DB 55774 GAGATTTTGAATTCGTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 55715
QY 361 CAGAAACAGGTTTAAACATTACGTTCCGAGACAGCTGCTATTGAGCGGAGGACCTTT 420
DB 55714 CAAAGCAAGGTTTACATTACATTGGAATACATTTTGTAAAGTGTGCAACATTT 55655
QY 421 GCTATCATGTTGTTTACTTAAAT 447
DB 55654 GCAATCATGAATATTATTACTTATTT 55628
```

Search completed: February 27, 2006, 11:11:04
Job time : 2576.57 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:36:56 ; Search time 2522.91 Seconds
(without alignments)
8289.569 Million cell updates/sec

Title: US-10-789-164-2

Sequence: 447
1 atgttcgacacgaattata.....tgccttcttactactaact 447

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: gb_esc1:*

2: gb_esc2:*

3: gb_esc3:*

4: gb_hic:*

5: gb_esc4:*

6: gb_esc5:*

7: gb_esc6:*

8: gb_esc7:*

9: gb_gsa1:*

10: gb_gsa2:*

11: gb_gsa3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Query Length	DB ID	Description
1	42.8	9.6	296	1	AM876109 CML-PT001
2	40.8	9.1	973	10	CL235356 ZMMBB057
3	39.8	8.9	778	10	AG490873 Mus muscu
4	39.6	8.9	258	10	CG907847 ZMMBB052
5	39.6	8.9	905	9	BZ698247 PUBLV12TD
6	39.4	8.8	646	9	BH530334 BOCQ245TR
7	39.4	8.8	666	10	CL941698 CA_ABA005
8	39.4	8.8	771	9	BZ601245 WHACX08TF
9	39.4	8.7	559	9	BH783640 fmbd013f0
10	38.6	8.6	421	1	AUI75308 AUI75308
11	38.6	8.6	568	6	CD323072 STRPUS37
12	38.4	8.6	764	8	DN041664 JGI CABAI
13	38.4	8.6	887	8	DR732535 FGAS07845
14	38.4	8.6	1007	9	BH135121 ENTNZ59TR
15	38.4	8.5	534	9	AQ765914 HS_5421_A
16	38.4	8.5	779	10	CL150443 104_332_1
17	37.8	8.5	577	9	AZ399037 1M0164N09
18	37.6	8.4	667	7	CK654217 AGENCOURT
19	37.4	8.4	487	6	CF811624 NA838 CDN
20	37.4	8.4	627	10	AG921347 Drosophila
21	37.4	8.4	849	10	DU002865 289737 To
22	37.4	8.4	858	5	BX440759 BX440759

C 23	37.2	8.3	1032	7	CN645534 ILUMIGEN
C 24	37	8.3	446	10	CZ073605 Ba006
C 25	37	8.3	490	9	BH267282 CH230-138
C 26	37	8.3	694	9	BH355163 CH230-81P
C 27	36.8	8.2	397	11	CNS03XCK
C 28	36.8	8.2	683	9	BH932950
C 29	36.6	8.2	1170	11	CNS06QF9
C 30	36.6	8.2	641	9	AZ835046
C 31	36.6	8.2	726	9	AZ317662 1M0036008
C 32	36.6	8.2	808	9	BH096957 RPCI-24-2
C 33	36.6	8.2	1712	4	CNS0R4QY
C 34	36.4	8.1	619	11	DE054014
C 35	36.2	8.1	618	11	DE053326
C 36	36.2	8.1	673	1	AV653326
C 37	36.2	8.1	750	10	CNS011TD
C 38	36.2	8.1	961	11	CNS06EOW
C 39	36.2	8.1	1030	10	CL056667
C 40	36	8.1	503	9	AQ534563
C 41	36	8.1	694	6	CD904177
C 42	36	8.1	738	9	AQ210826
C 43	36	8.1	799	5	BX446786
C 44	36	8.1	932	5	BQ669779
C 45	36	8.1	1101	10	CNS0036G

ALIGNMENTS

RESULT 1
AM876109 296 bp mRNA linear EST 22-MAY-2000
LOCUS CML-PT0016-231299-075-d08 PT0016 Homo sapiens CDNA, mRNA sequence.
DEFINITION AM876109.1 GI:8014132
ACCESSION AM876109.1 GI:8014132
VERSION
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Homo sapiens
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.
1 (bases 1 to 296)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagal,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Coستا,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., de Oliveira,P.S., Bucher,P., Jongeneel,C.V.,
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
10737800
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=cl2-CML-PT0016-231
299-075-d08&cl=1999-12-23&cl=1)
Seq primer: puc 18 forward
High quality sequence start: 16
High quality sequence stop: 296.
Location/Qualifiers
1..296
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="PT0016"

Qy 176 TTATCGTAATACCGTTTATCCAAATTATGATTTTGTACGACGACGAAATTCGCTG 235
 Db 567 TTAGCGTGGTTAAAGGTTTATCATTTGTTGTTTCTAA-----AAAAATAATAC 620
 Qy 236 CCATCGTATACAGGAGATCGCCAAAATGCGTTGATTTTCTATACCGGATG 292
 Db 621 CTCTCTGGTTGGGAGATTCATTAAGATAGCTCTTTTCTTCTTCTACCTGGGTG 677

RESULT 4

CG907847 259 bp DNA linear GSS 09-DEC-2003
 LOCUS ZMMBB0521G12r ZMMBB (HindIII) Zea mays genomic clone
 DEFINITION ZMMBB0521G12 3', genomic survey sequence.
 CG907847
 ACCESSION CG907847
 VERSION CG907847.1 GI:39607114
 KEYWORDS GSS.

SOURCE

ORGANISM Zea mays
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoideae; Andropogoneae; Zea.
 1 (bases 1 to 258)
 Bhatti,A.K., Young,S., Kavchok,S., Keizer,G., Bronzino,A.C.,
 Zohovetz,V., Fuks,G., Yu,Y., Wang,R. and Messing,J.
 Sequencing of the maize genome at PGR (2003c)
 Unpublished (2003)
 CONTACT: Bhatti,A.K.
 DR.Joachim Messing's lab
 The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers
 University
 190 Frelinghuysen Road, Piscataway, NJ 08854, USA
 Tel: 732 445 3801
 Fax: 732 445 5735
 Email: bhatti@waksman.rutgers.edu
 Seq primer: Spe
 Class: BAC ends
 High quality sequence start: 112.
 Location/Qualifiers

FEATURES

1..258
 /organism="Zea mays"
 /mol_type="genomic DNA"
 /cultiivar="B73"
 /db_xref="taxon:4577"
 /clone="ZMMBB0521G12"
 /lab_host="E. coli DH10B"
 /clone_lib="ZMMBB (HindIII)"
 /note="Vector: pCUG1; Site_1: HindIII; Site_2: HindIII"
 ORIGIN

ORIGIN

Query Match 8.9%; Score 39.6; DB 10; Length 258;
 Best Local Similarity 63.8%; Pred. No. 1.9;
 Matches 60; Conservative 0; Mismatches 34; Indels 0; Gaps 0;
 Qy 136 TTTATTTTACTGTTGTTGCTAGAGCTTCTCATTTATTCGTAATACGTTTA 195
 Db 129 TTTGTTTATATGTTTATATAGAAATCATGCTAATATATCATCATGAATGAGAGAGA 188
 Qy 196 TCCAAATTTATGATTTTGTACGACGACGAGAAAT 229
 Db 189 AGTTGAATTATGAACCTAGCCCTGAGCAGAAAT 222

RESULT 5

BZ698247/c 905 bp DNA linear GSS 19-FEB-2003
 LOCUS PUBLV12TD ZM 0.6_1.0 KB Zea mays genomic clone ZMMBTA084B24,
 DEFINITION genomic survey sequence.
 BZ698247
 ACCESSION BZ698247.1 GI:26418094
 VERSION BZ698247.1
 KEYWORDS GSS.
 SOURCE Zea mays

ORGANISM

Zea mays
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoideae; Andropogoneae; Zea.
 1 (bases 1 to 905)
 WhiteLaw,C.A., Quackenbush,J., Van Aken,S., Uteirback,T.,
 Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
 Bennezen,J.
 Maize Genomics Consortium
 Unpublished (2003)
 CONTACT: Cathy WhiteLaw
 TIGR

REFERENCE

AUTHORS
 TITLE
 JOURNAL
 COMMENT

FEATURES

source 1...905
 /organism="Zea mays"
 /mol_type="genomic DNA"
 /strain="B73"
 /db_xref="taxon:4577"
 /clone="ZMMBTA084B24"
 /clone_lib="ZM 0.6_1.0 KB"
 /note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
 cot selected genomic DNA library"
 ORIGIN

Query Match 8.9%; Score 39.6; DB 9; Length 905;
 Best Local Similarity 63.8%; Pred. No. 2.3;
 Matches 60; Conservative 0; Mismatches 34; Indels 0; Gaps 0;
 Qy 136 TTTATTTTACTGTTGTTGCTAGAGCTTCTCATTTATTCGTAATACGTTTA 195
 Db 770 TTTGTTTATATGTTTATATAGAAATCATGCTAATATATCATCATGAATGAGAGAGA 711
 Qy 196 TCCAAATTTATGATTTTGTACGACGACGAGAAAT 229
 Db 710 AGTTGAATTATGAACCTAGCCCTGAGCAGAAAT 677

RESULT 6

BH530334 646 bp DNA linear GSS 14-DEC-2001
 LOCUS BOGQ245TR BOGQ Brassica oleracea genomic clone BOGQ245, genomic
 survey sequence.
 BH530334
 ACCESSION BH530334.1 GI:117446779
 VERSION BH530334.1
 KEYWORDS GSS.
 ORGANISM Brassica oleracea
 Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
 1 (bases 1 to 646)
 Ayele,M., Haas,B.J., Kumar,N., Wu,H., Xiao,Y., Van Aken,S.,
 Uteirback,T.R., Mortman,J.R., White,O.R. and Town,C.D.
 Whole genome shotgun sequencing of Brassica oleracea and its
 application to gene discovery and annotation in Arabidopsis
 Genome Res. 15 (4), 487-495 (2005)
 15805490
 Other GSSs: BOGQ245TR
 CONTACT: Chris Town
 TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TR
 Class: sheared ends.

FEATURES
source Location/Qualifiers
1..666
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="T01000D93"
/db_xref="taxon:3712"
/clone_lib="BOGQ245"
/note="Vector: pHD108 T1 phage resistant"
genomic DNA inserted into pHD108 using BamHI linkers"

ORIGIN

Query Match 8.8%; Score 39.4; DB 9; Length 646;
Best Local Similarity 57.9%; Pred. No. 2.5;
Matches 70; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

OY 2 TGTTCGATCAGTATTATACATCGCAGTATTGTTAGTACTACTGTTTATTTTGG 61
DB 472 TGAATGATGATTTAGGTCATCGAGTTATCTTACTGTGGAGATCTGTTTTCG 531
OY 62 CGAATAAAGAGGATCGGCGGAGAGCTTGTGTACCGGATATTAGACTGTGT 121
DB 532 ACGGAAAAAGGAGAGTGGCGTGAAGATCATCTTCCGGGTTAGGTGACAGCAT 591
OY 122 T 122
DB 592 T 592

RESULT 7
LOCUS CL941698 666 bp DNA linear GSS 14-SEP-2004
DEFINITION OA_ABA0054107.f OA_ABA Oryza australiensis genomic clone
ACCESSION CL941698
VERSION CL941698.1 GI:52073279
KEYWORDS GSS.
SOURCE Oryza australiensis
ORGANISM Oryza australiensis
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jecy,R.,
Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
OMP Project
Unpublished (2004)
Contact: Rod A. Wing
Arizona Genomics Institute
University of Arizona
Forbes Building Room 303, Tucson, AZ 85721-0036, USA
Tel: 520 626 9595
Fax: 520 621 1259
Email: rwing@genome.arizona.edu
PCR Primers
FORWARD: TAA TAC GAC TCA CTA TAG GG
BACKWARD: CAC TCA TTA GGC ACC CCA
Plate: 0054 row: I column: 07
Seq primer: TAA TAC GAC TCA CTA TAG GG
Class: BAC ends.

TITLE
JOURNAL
COMMENT

FEATURES
source Location/Qualifiers
1..666
/organism="Oryza australiensis"
/mol_type="genomic DNA"
/db_xref="taxon:4532"
/clone_lib="OA_ABA0054107"
/issue_type="Young leaves"
/lab_host="DH108 T1 phage resistant"
/clone_lib="OA_ABA"
/note="Vector: pAGIBAC1, Site_1: HindIII, Site_2: HindIII"

ORIGIN

Query Match 8.8%; Score 39.4; DB 10; Length 666;

Best Local Similarity 55.5%; Pred. No. 2.5;
Matches 76; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

OY 169 ACTATATTTATCGGAATACGGTTTATCCAAATTTATGATTTTGTACGACGAGAAA 228
DB 491 AATATTTTATGATAAATAAGTTTGTGAATATATCATTTATTAATAGATATAATA 550
OY 229 TTCGTCGATCGATGATACAGGATCGCTTAATAATCGCGTTGATTTTCTATACCG 288
DB 551 TTTCTTCCCTTCAATATAGAAATCTTGTGAATAATGATTTTGTATTAATTTTTC 610
OY 289 ATTGACATTTGAAT 305
DB 611 TTTTATGAATTCAAAT 627

RESULT 8
LOCUS BZ601245 771 bp DNA linear GSS 08-JUN-2003
DEFINITION WHACX08TF Human MCF7 breast cancer cell line library (MCF7_1) Homo
sapiens genomic clone MCF7_1-18B16, genomic survey sequence.
ACCESSION BZ601245
VERSION BZ601245.1 GI:31509707
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1 (bases 1 to 771)
Volk,S., Zhao,S., Chin,K., Brebner,J.H., Herndon,D.R., Tao,Q.,
Kowbel,D., Huang,G., Lapuk,A., Kuo,W.-L., Magrane,G., de Jong,P.,
Gray,J.W. and Collins,C.
End-sequence profiling: Sequence-based analysis of aberrant genomes
Proc. Natl. Acad. Sci. U.S.A. 100 (13), 7696-7701 (2003)
1278976
Contact: Volk SV
Colin Collins' lab
UCSF Comprehensive Cancer Center
UCSF Box 0808, San Francisco, CA 94143-0808, USA
Tel: 415 502 7066
Fax: 415 502 5665
Email: svolk@cc.ucsf.edu
This clone is available from Amplicon Express
http://www.genomex.com
Class: BAC ends.

TITLE
JOURNAL
COMMENT

FEATURES
source Location/Qualifiers
1..771
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone_lib="MCF7_1-18B16"
/sex="female"
/clone_lib="Human MCF7 breast cancer cell line library
(MCF7_1)"
/note="Vector: pECBAC1, Site_1: HindIII; This library was
constructed from MCF7 breast cancer cell line by Amplicon
Express (http://www.genomex.com) using their standard
procedure."

ORIGIN
Query Match 8.8%; Score 39.4; DB 9; Length 771;
Best Local Similarity 47.1%; Pred. No. 2.6;
Matches 121; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

OY 80 TCGCGGAGACTGTTGTACCGGATATTAGACTGTGTTATATCAGCGGCTTTTA 139
DB 75 TGTATCTAGCGCTATATATATATAAGATCAGACTTTTATTAATTAACGCTTTT 134
OY 140 TTTACTGTTTTCCTAGTACGCTTGTCTCACTTATGTTATCGTAATAGCTTTATCA 199
DB 135 GTCAATTTCTTATTTGCTATTTTGAATAATTAATTTTGTGACAGAGCTTTGTA 194

QY 200 AATTATGATTTTGTACGAGCAGAAAATTCCTGCATGCTATACAGGATCTGCC 259
 DB 195 GTTTTCCTTTGATTTCTAAATACACTTTCCTTTCTATTTGTAAGATTAATGTTTC 254
 QY 260 TAAATATGCGGTTGATTTTCTATACCCGATGTACCAATTTGAATTCAGAAATTCGAG 319
 DB 255 TTATATTCACATTTTAATTCCTTTGGAGATTTAAGATTTATTCACAATAACTACTT 314
 QY 320 GAATCGGATCATCTGTG 336
 DB 315 TTCTCTCCATATTTGTTG 331
 RESULT 9
 BH783640/c 569 bp DNA linear GSS 28-MAR-2002
 LOCUS fzm013f008g12k0 fzm013f008g12 5', genomic survey sequence.
 DEFINITION BH783640
 ACCESSION BH783640.1 GI:19787191
 VERSION GSS.
 KEYWORDS Zea mays
 SOURCE Zea mays
 ORGANISM Zea mays
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoideae; Andropogonae; Zea.
 1 (bases 1 to 569)
 Budiman, M.A., Freese, R.G., Bedell, J.A., Nunberg, A.N. and Lakey, N.D.
 Genethresher methylation filtered genomic sequences from maize
 Unpublished (2002)
 CONTACT: Bedell JA
 ORION Genomics, LLC
 4041 Forest Park Ave, St. Louis, MO 63108, USA
 Tel: 314 615 6979
 Fax: 314 615 5975
 Email: jbedell@oriongenomics.com
 Plate: fzm013f008 row: 9 column: 12
 Seq primer: SK reverse
 Class: methylation filtered
 High quality sequence stop: 569.
 Location/Qualifiers
 1..569
 /organism="Zea mays"
 /mol_type="genomic DNA"
 /cultivar="Inbred line Mo17"
 /db_xref="taxon:4577"
 /clone="fzm013f008g12"
 /clone_lib="fzm013f008g12 filtered library"
 /note="Organ: Leaf; Vector: pBSK(-); Site: 1: HincII; DNA
 prepared from purified nuclei was randomly sheared,
 end-repaired, size fractionated to enrich for the 0.5 to
 5 kb fraction, ligated into HincII-digested pBSK(-)
 vector and electroporated into E. coli cells."

ORIGIN

Query Match 8.7%; Score 39; DB 9; Length 569;
 Best Local Similarity 51.4%; Pred. No. 3.2;
 Matches 90; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 40 GATACATCAGTTAATTTTGGCGAAAAACAGGATCTGCGCGAGACATTTGTA 99
 DB 272 GAACATACAGGTTTACCTGTAATAGAACCATAGAGGCCCTGATAGATCTTGTGT 213
 QY 100 CCGGATATTTAGACTGTGTTTAACACCCGCTCTTATTTACTGTTTGTCTAGTG 159
 DB 212 ACAGATTCATVAGAGGTTAAGACATTTTGACCTACGCTTACAGGTTGTCTACTT 153
 QY 160 AGCTGCTCACTTATGTTATCGTAATACGTTTATCCAAATTTATGATTTGT 214
 DB 152 TGCCTTCAGCTTTTGTGTGTCCCGACGCTTAGCTGAATGTCCTTTT 98
 RESULT 10

AUI75308 421 bp mRNA linear EST 11-JUN-2003
 LOCUS AUI75308
 DEFINITION AUI75308 Marsupenaeus japonicus adult Marsupenaeus japonicus cDNA
 clone Puh366 forward similar to Leishmania major LmJ393/298A
 Leishmania major promastigote full length cDNA library from early,
 mRNA sequence.
 ACCESSION AUI75308
 VERSION AUI75308.1 GI:14621715
 KEYWORDS EST.
 SOURCE Marsupenaeus japonicus
 ORGANISM Marsupenaeus japonicus
 Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
 Eumalacostraca; Eucarida; Decapoda; Dendrobranchiata; Penaeidea;
 Penaeidae; Marsupenaeus.
 1 (bases 1 to 421)
 Rojtimakorn, J., Hirano, I., Itami, T., Takahashi, Y. and Aoki, T.
 Gene expression in haemocytes of kuruma prawn, Penaeus japonicus,
 in response to infection with WSSV by EST approach
 Fish Shellfish Immunol. 13 (1), 69-83 (2002)
 12201653
 CONTACT: Takashi Aoki
 Aquatic Biosciences
 Tokyo University of Fisheries
 Komae 4-5-7, Tokyo, Minato-ku 108-8477, Japan
 Tel.: 81-3-5463-0689
 Fax: 81-3-5463-0690
 Email: aoki@tokyo-u-fish.ac.jp.
 Location/Qualifiers
 1..421
 /organism="Marsupenaeus japonicus"
 /mol_type="mRNA"
 /db_xref="taxon:27405"
 /clone="Puh366"
 /cell_type="hemocytes"
 /dev_stage="adult"
 /clone_lib="Marsupenaeus japonicus adult"
 /note="Vector: lambda Zap II"

ORIGIN

Query Match 8.6%; Score 38.6; DB 1; Length 421;
 Best Local Similarity 36.7%; Pred. No. 4;
 Matches 81; Conservative 35; Mismatches 105; Indels 0; Gaps 0;

QY 134 TCTTATTTTACTGTTTGTCTAGTAGAGCTGCTCACTATGTAATGTAATACGGTT 193
 DB 1 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 60
 QY 194 TATCCAAATTTATGATTTTGTACGAGCAGCAAAAATTCGCTGCATGCTATACAGGA 253
 DB 61 YYYCAATTAATAAKTWTGCAAGCAGCMTTMMAGCASCAGCTTAAATTTKAGYTTKKYCY 120
 QY 254 TCGTCTTAAATAATCGGTTGATTTTCTATVATCCAGATGTATCAATTTGAATTCGAGAA 313
 DB 121 YCYCCAKTKGSGSSYTTGGWTTTMMYTTTMYTTTTCATTAASGKTYCYHAMWCC 180
 QY 314 TTGAGGAATCGGATCATCTGCGCAGGTTTAATGCCAT 354
 DB 181 WTGKGGAAMAGSCYTTTYYTTTYYTTTYYTTTCCMAAY 221
 RESULT 11
 CD323072 568 bp mRNA linear EST 17-SEP-2003
 LOCUS CD323072
 DEFINITION StrPu537.001434 Sea urchin embryo 20th blastula stage cDNA library
 MPMG537 Strongylocentrotus purpuratus cDNA clone
 CALDB53701524.MPI_537_24015 5', mRNA sequence.
 ACCESSION CD323072
 VERSION CD323072.1 GI:34795133
 KEYWORDS EST.
 SOURCE Strongylocentrotus purpuratus
 ORGANISM Strongylocentrotus purpuratus
 Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
 Echinoidea; Echinoidea; Echinacea; Echinoida;

REFERENCE	TITLE	JOURNAL	COMMENT
Strongylocentrotidae; Strongylocentrotus. 1 (bases 1 to 568)	Poustka,A.J., Groth, D., Hennig, S., Thamm, S., Cameron, A., Beck, A., Reinhardt, R., Herwig, R., Panopoulou, G. and Lehnrich, H.	Genetec, annotation, evolutionary analysis, and database integration of 20,000 unique sea urchin EST clusters	
	Genome Res. 13 (12), 2736-2746 (2003)		
	Contact: Poustka AJ		
	laboraty 145, dept. Lehnrich		
	Max-Planck-Institut fuer Molekulare Genetik		
	Innestr.63-73, D-14195 Berlin, Germany		
	Tel.: +49 30 8413 1235		
	Fax: +49 30 8413 1128		
	Email: poustka@molgen.mpg.de		
	The library was characterised by oligonucleotide fingerprinting (ONP) to reduce sequencing redundancy. According to the ONP procedure, clones that display the same hybridisation matrix with a battery of 200 8mer oligonucleotides are grouped into clusters. One clone per ONP cluster is selected for sequencing. The size of each cluster is an indicator of the frequency of a transcript in the analysed library. The cluster size as well as the coordinates of the other clones assigned to the same ONP cluster as the clone from which the above EST is generated is available at the sea urchin project web site at: http://www.molgen.mpg.de/ag_seaurchin/ . CDNA Database of the German Human Genome Project (http://www.tzpd.de)		
	PCR Primers		
	FORWARD: 5' CCCACGCTTACACTTATGCTTCGGGCTG 3' (M13SP) 5'-seq		
	BACKWARD: 5' GCATTACGCGAGCGAGCAAGGGAGATGTG 3' (M13FS) 3'-seq		
	Seq primer: 5'-CCGATCCGGAATTCGGGCT-3' pSPORT3/86		
	High quality sequence stop:568.		

FEATURES		source		location/Qualifiers	
1..568				organism="Strongylocentrotus purpuratus"	
				mol_type="mRNA"	
				/db_xref="taxon:7668"	
				/clone="CALR53701524;MPI_537_24015"	
				/tissue_type="whole embryo"	
				/dev_stage="embryonic 20hr"	
				/lab_host="E.coli, XL1 blue"	
				/clone_id="Sea urchin embryo 20hr blastula stage cDNA library MPW6537"	
				/note="Vector: pSport1; Site_1: NotI; Site_2: SalI; Random primed and directionally cloned in pSport1 vector using a NotI (5'-pGATAGTTCTGATCGCAGCGGCGGCC (T)15-3' and a SalI 5'-TCGACCACGCGCTCCG-3' adapters (Gibco BRL)"	
ORIGIN					
Query Match	8.6%	Score 38.6;	DB 6;	Length 568;	
Best Local Similarity	51.4%	Pred. No. 4.2;	84;	Indels 0;	Gaps 0;
Matches	89;	Conservative 0;	Mismatches		
Qy	133	GTCTTATTTACTGTTTGTGCTAGTGAAGCTTGCTCACTTATGTTATCGTAAATACGGT	192		
Db	8	GTCCGCTGTTAGTTGTGCTCAAGTTCCTGTTTACTATGTTCTGTCTTCATCTGT	67		
Qy	193	TTATTCCAATTATGATTTTGTAGCAGCAGCAGAAATTCGCTGCATCCTGATACAGG	252		
Db	68	AGATTATTCATCAGATATATATATTATTCAGAGATTATTTGATATCTCTTATATCTAGG	127		
Qy	253	ATCGCTCAAAATCGCGTTGATTTCTATACCCGATTTGATACATTGGAAT	305		
Db	128	TTTGTATATATTTCAATTTTTTCTTTTGTGTTGTCGATTTGGTTTTATGAT	180		
RESULT 12					
DN041664/c					
LOCUS					
DEFINITION JGI CABAL136.rev NIH XGC tropickd1 Xenopus tropicalis cDNA clone					
IMAGE:7746695 3', mRNA sequence.					
ACCESSION DN041664					
VERSION DN041664.1					
KEYWORDS GI:59247540					
EST.					

SOURCE	ORGANISM	REFERENCE	TITLE	JOURNAL	COMMENT
Xenopus tropicalis (western clawed frog)	Xenopus tropicalis	1 (bases 1 to 764)	Richardson, P., Lucas, S., Rokhsar, D., Dettler, J. C., Ng, D. C., Brokselien, P. and Lindquist, E. A.	DOE Joint Genome Institute Xenopus tropicalis EST project	Unpublished (2004)
Other ESTs: UGI, CABAI336, fwd	Contact: Lindquist, E. A., Richardson, P.	DOE Joint Genome Institute	2800 Mitchell Drive, Walnut Creek, CA 94598, USA	Tel: 925 296 5600	Fax: 925 296 5710
Email: cdnaajgl@psf.org	Tissue Procurement: Robert M. Grainger	CDNA Library Preparation: Bruce Blumberg Laboratory, University of California, Irvine	DNA Sequencing: DOE Joint Genome Institute	Clone Distribution: I.M.A.G.E. Consortium/LLNL	http://image.llnl.gov
Naming Conventions: EST name is generated by the concatenation of the UGI Clone id and the direction of sequencing. The suffix '.rev' indicates a reverse sequencing read of the insert. It does not necessarily reflect the orientation of the insert.	Plate: CABA 0013 row: P column: 21	High quality sequence stop: 756.	location/Qualifiers	1..764	

```

/organism="xenopus tropicalis"
/mol_type="mRNA"
/strain="N6 (Nigerian 6th generation inbred)"
/db_xref="taxon:8364"
/clone="IMAGE:7746695"
/tissue="kidney"
/dev_stage="adult"
/lab_host="ElectroMAX DH10B T1 Phage Resistant cells"
/clone_id="NH_XGC_tropKid1"
/notes="Vector: pCS107; Site 1: EcoRI; site 2: XhoI; The library was prepared from 5 ug of poly A+ RNA by oligo-dT priming"
(5'-GACAGAGAGAGAGAGAGACTGCTCAGCTTTTTTTTTTTTT-3')
and SacraScript reverse transcriptase. After ligation of
EcoRI adaptlers (5'-AATTCGGACAGG-3') followed by kinsing
adapters and by XhoI digestion, the cDNA was size selected
by chromatography on Sepharose CL-2B columns and fractions
containing cDNAs larger than 1000 bp were ligated into
EcoRI/XhoI-digested pCS107. Reference for library
construction: Current genome 4, 635-644. Library
constructed by Michelle Tabb and Bruce Blumberg (Dept of
Developmental and Cell Biology, University of California,
Irvine)."

ORIGIN

Query Match      8.6%; Score 38.4; DB 8; Length 764;
Best Local Similarity 58.9%; Pred. No. 5;
Matches 66; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY      104 GATATTAGGACTGTGTGTTTAAATCAGCCGCTTTATTATTACTCTGTTTCTAGTACCT 163
      |||
      |||
      |||
DB      400 GAGTTTATTGTTTCTGTATACCCAGTCAGATATATATATATTGTTGAGTGACCT 341
      |||
      |||
      |||
QY      164 TGTTCACCTTAACTTATCGTGAAGAAGCGTTTATCCAAATTATGATTTGTA 215
      |||
      |||
      |||
DB      340 CTCTTATTAGTTACAGATTATATATATATACAGATTATATATATATA 289
      |||
      |||
      |||

RESULT 13
LOCUS      DR732535      887 bp      mRNA      linear      EST 18-JUN-2005
DEFINITION FGAS078455 Triticum aestivum FGAS: TaL65 Triticum aestivum CDNA,

```

ACCESSION	mRNA sequence.
VERSION	DR732535
KEYWORDS	DR732535.1 GI:70959192
SOURCE	EST.
ORGANISM	Triticum aestivum (bread wheat)
REFERENCE	Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae; Triticum. 1 (bases 1 to 887) Allard, F., Crosby, W.L., Danyluk, J., Endes, F., Frick, M., Gaudet, D., Gawsein, B., Graf, R., Gulick, P., Hrycan, L.D., Iacoché, A., Links, M.G., McCarthy, E.L., Monroy, A., Muzak, I., Nilsson, D., Penniket, C., Roach, J.L. and Sarhan, F. Functional Genomics of Abiotic Stress In Wheat and Canola Crops Unpublished (2003) Contact: Patrick Gulick Plant Molecular Biology Concordia University, Department of Biology 7141 Sherbrooke St. West, Montreal, Quebec H4B 1K6, Canada Tel.: 514 848 2424 Ext 3407 Fax: 514 848 2881 Email: pgulick@alcor.concordia.ca This sequence is the direct result of the Base calling software phred (default parameters). It is the raw base calls. To aid in the identification of the high quality insert the software Lucy (default parameters) has been run on this sequence. Lucy identified the region [123..277]. Plate: Talts40 row: K column: 22.
TITLE	Location/Qualifiers
JOURNAL	1. 887
COMMENT	/organism="Triticum aestivum" /mol_type="mRNA" /cultivar="wheat line PI 178383" /db_xref="taxon:4565" /lab_host="DH5 alpha" /clone_lib="Triticum aestivum FGAS: Talts5" /note="Organ: Crown; Vector: pGEM-T; SSH (suppression subtractive hybridization) cDNA library from genotype PI178383 cold hardened at 2 C for 21 days and 49 days (equal amount of cDNA pooled together before subtraction, tester) and subtracted against genotype Norstar cold hardened at 2 C for 1 day (24 h) (driver). Modified Smart cDNA (Clontech) priming and non-directional cloning"
FEATURES	
SOURCE	
ORIGIN	
Query Match	86%; Score 38.4; DB 8; Length 887;
Best Local Similarity	48.2%; Pred. No.5.1; Indels 0; Gaps 0;
Matches 108; Conservative	0; Mismatches 116; Indels 0; Gaps 0;
Oy	94 GTTGACCGGGATATTAGACTGTGTGTTATCAGCCGCTTATTATTACTGTTTGG 153
Db	
Oy	127 GTGGCGCGGCCGAGSTGGCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 186
Db	
Oy	154 CTAGTAGCTTGTCACCTTAATGTTATCGTGAATACGGTTATCCAATTATGATTGG 213
Db	
Oy	187 TTTTGT 246
Db	
Oy	214 TAGGACCGCAAAATTCGCTGCCTCATGCTAACAAGGATCGCTCAAATCGCGTTT 273
Db	
Oy	247 CACTATCTTATTGTTTTTTTTTTTTTTTTTTTACTACTGCGAGAGCTGTATGAGACGTGTT 306
Db	
Oy	274 GATTTCTATAACCGATTGTACCATTTGAATACGAGAATTTTCG 317
Db	
Oy	307 CAATTTTATATCGAATGTACCTCTCAAAATATCTGATTTTGG 350
Db	
RESULT 14	
BH135121/c	1007 bp DNA linear GSS 07-AUG-2001
LOCUS	ENTNZ55TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION	genomic, genomic survey sequence.
ACCENSION	BH135121

VERSION	GI:15094182
KEYWORDS	GSS.
SOURCE	Entamoeba histolytica
ORGANISM	Entamoeba histolytica
REFERENCE	Entamoeba histolytica
ATTNORS	1 (bases 1 to 1007)
TITLE	Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.
JOURNAL	Determination of clone end sequences from Entamoeba histolytica
COMMENT	HMI:IMSS sheared DNA library (2001)
	Unpublished (2001)
	Contact: Brendan J Loftus
	Department of Eukaryotic Genomics
	The Institute for Genomic Research
	912 Medical Center Dr., Rockville, MD 20850, USA
	Tel: 301 838 0208
	Fax: 301 838 3543
	Email: b.loftus@tigr.org
	Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
	DNA library
	Seq primer: M13-Reverse
	Class: Shotgun
	High quality sequence start: 2
	High quality sequence stop: 418.
	Location/Qualifiers
FEATURES	1. 1007
source	"/organism="Entamoeba histolytica"
	"/mol_type="genomic DNA"
	"/accession="HMI:IMSS"
	"/db_xref="taxon:5759"
	"/clone_id="Entamoeba histolytica sheared DNA"
	"/notes="Vector: PHOS1; Site 1: Bst I; Constructed at The
	Institute for Genomic Research (TIGR), Rockville, MD.
	Genomic DNA isolated from broth cultures of E. histolytica
	using a method described by Clark and Diamond (Clark,
	C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
	method for isolate identification. Exp. Parasitol.
	77:450). The DNA was mechanically sheared to give a
	tight size distribution (~2 kb). The v + i method used for
	the library construction is described in detail in Smith,
	H.O., and Venter, J.C. (Making small insert libraries for
	whole genome shotgun sequencing projects. In Genome
	Sequencing: A Practical Approach, ed. M. Vaudin and B.
	Barell, Oxford University Press, 1999)."
ORIGIN	
Query Match	8.6%; Score 38.4; DB 9; Length 1007;
Best Local Similarity	48.6%; Pred. No. 5.2;
Matches 105; Conservative 0; Mismatches 111; Indels 0; Gaps 0;	
QY	12 AGATTATATACATGCGCAATATTTAGGTGATCTACTAGTTATTTTGGCGAAAAAC 71
DB	686 AAATGTGATTAATGAAATTTGGCACAATTAATTAATATATATATCTATTTCTTAAGCAATTA 627
QY	72 AGGATGCGTGGCGGACGAGACTGTGATACCGGATATTTAGGACTTGTTGTTATTCAGCC 131
DB	626 ATGAAATGACATTCATTCATTTTATATACCTGCGACAAAGTATTAAGACTTAAATGTT 567
QY	132 GGTCTTATATTTACTTGTGTTTGTAGTAGAGCTTCACTTATGTATGCTGAATACGG 191
DB	566 TTTTATTAATTAATGAGTTAATGAATGAGTTAGTAGGAGATATCGTTAATATCT 507
QY	192 TTTATCCAAATTTATGATTTTGTACGACGACGAAA 227
DB	506 ATAAATAAATTAATTAATTAATGTTGTAATTAATAA 471
RESULT 15	
LOCUS	A0765914/C 534 bp DNA linear GSS 28-JUL-1999
DEFINITION	HS 5421 A2 G05 T7A RPCI-11 Human Male BAC Library Homo sapiens
ACCESSION	genomic clone Plate=97 Col=10 Row=M, genomic survey sequence.
VERSION	A0765914 GI:5644030

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:34:16 ; Search time 282.194 Seconds
(without alignments)
10556.990 Million cell updates/sec

Title: US-10-789-164-2

Perfect score: 447
Sequence: 1 atgttcgacatcagattacta.....tgtttgttactactaatt 447

Scoring table: IDENTITY NUC
Gapop 10_0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq_21.*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	447	100.0	447	10	ADA44792 Bacillus
2	447	100.0	447	13	AD007642 B subtilis
3	447	100.0	447	13	ADV65733 B. subtilis
4	447	100.0	6536	10	ADA44796 Microbial
5	440.6	98.6	450	4	AAFG2256
6	440.6	98.6	3045	4	AAFG2254
7	289.2	64.7	450	6	ABK75634 Bacillus
8	189.4	42.4	453	4	AAH53465 S. epider
9	189.4	42.4	459	6	ABN92026 Staphyloc
10	189.4	42.4	459	13	ADSO1974 Staphyloc
11	182	40.7	3240	4	AAH53379 S. epider
12	65.2	14.6	480	9	ADB06111 Altiolococ
13	65.2	14.6	480	9	ADB06109 Altiolococ
14	65.2	14.6	110000	9	ADBI2064 Altiolococ
15	52.8	11.8	198	6	ABN92026 Staphyloc
16	45.6	10.2	3029	4	AAH54833 S. epider
17	43.2	9.7	2000	11	ACI37108 Rice stre
18	37.4	8.4	5822	6	ABL33096 Human imm
19	36.2	8.1	26997	4	AAS46748 Tumour su

20	36.2	8.1	110000	10	ADF77343_06	Continuation (7 of
21	36	8.1	7195	4	AAS45325	Aas45325 Chemical
22	36	8.1	7195	6	ABK28166	ABK28166 DNA trans
23	35.2	7.9	2100	6	ABK72830	ABK72930 Bacillus
24	35.2	7.9	4078	6	AAD29896	Aad29896 B. lichen
25	35	7.8	1347	8	ACNA9970	Acna9970 Prokaryot
26	35	7.8	2116	12	ADN12148	Adn12148 Interleuk
27	35	7.8	2116	13	ACF87528	Acf87528 Human SIR
28	35	7.8	2116	14	AD226514	Ad226514 Human IL7
29	35	7.8	11839	4	ABL04940	Ab104940 Human rep
30	35	7.8	11839	4	ABL97834	Ab197834 Human tes
31	35	7.8	11839	8	AB274494	Ab274494 Secreted
32	35	7.8	11839	8	ADA44503	Ada44503 Human sec
33	34.8	7.8	2000	11	ACL37886	Ac137886 Rice stre
34	34.8	7.8	2000	11	ACL35887	Ac135887 Rice stre
35	34.4	7.7	534	4	AAF24848	Aaf24848 Degenerat
36	34.4	7.7	534	6	AD42512	Ad42512 Human zal
37	34.4	7.7	534	9	ACH00107	Ach00107 Dehgenera
38	34.4	7.7	5769	6	ABK40019	Abk40019 Human che
39	34.4	7.7	51259	2	AAK83007	Aax83007 Partial m
40	34.2	7.7	1344	10	ABX07028	Abx07028 S. pneumo
41	34.2	7.7	5273	6	ABL32874	Ab132874 Human imm
42	34.2	7.7	6052	6	ABL32417	Ab132417 Human imm
43	34.2	7.7	12666	2	AAV52270	Aav52270 Streptoco
44	34.2	7.7	110000	10	ABS56454_11	Continuation (12 o
45	34.2	7.7	277616	13	ABD32602	Abd32602 Human can

ALIGNMENTS

RESULT 1	ADA44792	ADA44792 standard; DNA; 447 BP.
ID	ADA44792	standard; DNA; 447 BP.
XX	ADA44792;	
DT	20-NOV-2003	(first entry)
XX	Bacillus subtilis poly-gamma-glutamate synthetase complex pgsC gene.	
XX	Cell surface; expression vector; microbial;	
KW	poly-gamma-glutamate synthetase; pgsBCA complex; surface expression;	
KW	Gram-positive bacterium; Gram-negative bacterium; enzyme; antigen;	
KW	antibody; attachment protein; adsorption protein; vaccine; pgsC; gene;	
OS	Bacillus subtilis; variety chungkookjang.	
PN	WO2003014360-A1.	
XX	20-FEB-2003.	
XX	09-AUG-2002; 2002WO-KR001522.	
XX	10-AUG-2001; 2001KR-00048373.	
PA	(BIOL-) BIOLEADERS CORP.	
PA	(MDMD-) MD LAB CO LTD.	
XX	Sung M, Hong S, Lee J, Jung C, Kim C, Soda K, Ashiuchi M;	
XX	WPI; 2003-256589/25.	
PT	New expression vector containing gene(s) that encode a poly-gamma-	
PT	glutamate synthetase complex, useful for producing proteins (e.g.	
PT	vaccines or enzymes) on the microbial surface of Gram-positive and/or	
XX	Gram-negative bacteria.	
PS	Claim 3; Page 11; 122pp; English.	
XX	The invention relates to a vector for expression of a target protein on a	
CC	microbial cell surface. The vector of the invention comprises either one	

or more than two genes of the *Bacillus subtilis* poly-gamma-glutamate synthetase complex (pgsBCA) to facilitate microbial surface expression of the target protein. The pgsBCA gene complex comprises the pgsB, pgsB and pgsA genes and is normally expressed in the outer membrane of *Bacillus subtilis*. The vector can be transformed into either Gram-positive or Gram-negative bacteria (e.g., *Escherichia coli*), and can be used for the surface expression of various proteins of interest such as enzymes, antigens, antibodies, attachment proteins or adsorption proteins. Proteins recombinantly produced using the vector of the invention can be used as, for example, vaccines or enzymes. The present sequence represents the *Bacillus subtilis* pgsBCA complex gene pgsC, which is specifically claimed for use in the vector of the invention.

Sequence 447 BP; 116 A; 84 C; 94 G; 153 T; 0 U; 0 Other;

Query Match 100.0%; Score 447; DB 10; Length 447;

Best Local Similarity 100.0%; Pred. No. 2.7e-126;

Matches 447; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 ATGTTCCGATCAGATTATATACATCGCACTAATTTAGGAGTACTACTAGTTAATTTT 60
DB 1 ATGTTCCGATCAGATTATATACATCGCACTAATTTAGGAGTACTACTAGTTAATTTT 60
OY 61 GCGGAAAAACAGGGATCGTCCGCGAGACTTGTGTACCGGATATTAGACTTGTG 120
DB 61 GCGGAAAAACAGGGATCGTCCGCGAGACTTGTGTACCGGATATTAGACTTGTG 120
OY 121 TTTATACGCGGCTTTATTTTATTTGTTGCTAGAGAGCTTCTACTATATATC 180
DB 121 TTTATACGCGGCTTTATTTTATTTGTTGCTAGAGAGCTTCTACTATATATC 180
OY 121 TTTATACGCGGCTTTATTTTATTTGTTGCTAGAGAGCTTCTACTATATATC 180
DB 121 TTTATACGCGGCTTTATTTTATTTGTTGCTAGAGAGCTTCTACTATATATC 180
OY 181 GTGAATACGGTTTATCCAAATTTATGATTTGTACGAGCAGAAAAATTCGCTGCATG 240
DB 181 GTGAATACGGTTTATCCAAATTTATGATTTGTACGAGCAGAAAAATTCGCTGCATG 240
OY 241 CTGATTAACAGGATCGTCTAAATAATCGCTTTGATTTTCTATACCGCATTTGATCATTT 300
DB 241 CTGATTAACAGGATCGTCTAAATAATCGCTTTGATTTTCTATACCGCATTTGATCATTT 300
OY 301 GAAATCGCAGAAATTCGAGGAATCGGCATCATCGTCCAGGTTTATTTGCCAATTCATTT 360
DB 301 GAAATCGCAGAAATTCGAGGAATCGGCATCATCGTCCAGGTTTATTTGCCAATTCATTT 360
OY 361 CAGAAACAGAGTTTAAACATTACGTTTGGAAAGCAGCGTCTATTGAGGAGGAGCAGCTTT 420
DB 361 CAGAAACAGAGTTTAAACATTACGTTTGGAAAGCAGCGTCTATTGAGGAGGAGCAGCTTT 420
OY 421 GCTATCATGTTTGTCTTACTACTTAATT 447
DB 421 GCTATCATGTTTGTCTTACTACTTAATT 447
```

RESULT 2

ADO07642 ID ADO07642 standard; DNA; 447 BP.

ADO07642; ADO07642;

DT 15-JUL-2004 (first entry)

DE B subtilis poly-X-glutamate synthetase complex coding sequence pgsC.

XX ds; gene; enzyme; vaccine; cytosolic; pgsC; poly-X-glutamate synthetase;

KM human papilloma virus.

XX *Bacillus subtilis*.

OS WO2004035795-A1.

XX 29-APR-2004.

PF 17-OCT-2003; 2003WO-KR002163.

PR 17-OCT-2002; 2002KR-00063378.

XX

PA (BIOL-) BIOLEADERS CORP.

PA (KORE-) KOREA RES INST BIOSCIENCE & BIOTECHNOLOG.

XX

PI Sung M, Poo H, Lee J, Jung C, Hong S, Kim C, Park S, Pyo H;

XX WPI; 2004-348463/32.

DR

XX New vector containing pgs A-C genes encoding poly-gamma-glutamate

PT synthetase complex and an antigen protein gene of human papilloma virus,

PT useful in preparing vaccine for treating or preventing mucosal tumor,

PT e.g. cervical cancer.

XX

PS Disclosure; Page 61-62; 69pp; English.

XX

CC The present invention relates to a vector for preparing a vaccine which

CC contains one or more than two genes, i.e. pgs A-C encoding poly-X-

CC glutamate synthetase complex and an antigen protein gene of human

CC papilloma virus. The vector and microbes transformed with it are useful

CC in preparing vaccines for treating or preventing mucosal tumor, e.g.

CC cervical cancer. The present sequence is a *Bacillus subtilis* poly-X-

CC glutamate synthetase complex coding sequence.

XX

Sequence 447 BP; 116 A; 84 C; 94 G; 153 T; 0 U; 0 Other;

Query Match 100.0%; Score 447; DB 12; Length 447;

Best Local Similarity 100.0%; Pred. No. 2.7e-126;

Matches 447; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 ATGTTCCGATCAGATTATATACATCGCACTAATTTAGGAGTACTACTAGTTAATTTT 60
DB 1 ATGTTCCGATCAGATTATATACATCGCACTAATTTAGGAGTACTACTAGTTAATTTT 60
OY 61 GCGGAAAAACAGGGATCGTCCGCGAGACTTGTGTACCGGATATTAGACTTGTG 120
DB 61 GCGGAAAAACAGGGATCGTCCGCGAGACTTGTGTACCGGATATTAGACTTGTG 120
OY 121 TTTATACGCGGCTTTATTTTATTTGTTGCTAGAGAGCTTCTACTATATATC 180
DB 121 TTTATACGCGGCTTTATTTTATTTGTTGCTAGAGAGCTTCTACTATATATC 180
OY 121 TTTATACGCGGCTTTATTTTATTTGTTGCTAGAGAGCTTCTACTATATATC 180
DB 121 TTTATACGCGGCTTTATTTTATTTGTTGCTAGAGAGCTTCTACTATATATC 180
OY 181 GTGAATACGGTTTATCCAAATTTATGATTTGTACGAGCAGAAAAATTCGCTGCATG 240
DB 181 GTGAATACGGTTTATCCAAATTTATGATTTGTACGAGCAGAAAAATTCGCTGCATG 240
OY 241 CTGATTAACAGGATCGTCTAAATAATCGCTTTGATTTTCTATACCGCATTTGATCATTT 300
DB 241 CTGATTAACAGGATCGTCTAAATAATCGCTTTGATTTTCTATACCGCATTTGATCATTT 300
OY 301 GAAATCGCAGAAATTCGAGGAATCGGCATCATCGTCCAGGTTTATTTGCCAATTCATTT 360
DB 301 GAAATCGCAGAAATTCGAGGAATCGGCATCATCGTCCAGGTTTATTTGCCAATTCATTT 360
OY 361 CAGAAACAGAGTTTAAACATTACGTTTGGAAAGCAGCGTCTATTGAGGAGGAGCAGCTTT 420
DB 361 CAGAAACAGAGTTTAAACATTACGTTTGGAAAGCAGCGTCTATTGAGGAGGAGCAGCTTT 420
OY 421 GCTATCATGTTTGTCTTACTACTTAATT 447
DB 421 GCTATCATGTTTGTCTTACTACTTAATT 447
```

RESULT 3

ADV65733 ID ADV65733 standard; DNA; 447 BP.

ADV65733; ADV65733;

XX 10-FEB-2005 (first entry)

DE B. subtilis poly-gamma-glutamic acid synthase C gene SEQ ID NO:2.

XX

KW	des; poly-gamma-glutamic acid synthase C; pgsC; gene expression;
XX	antibiotic; antimicrobial; fungicide; cytostatic.
OS	Bacillus subtilis.
XX	
PN	KR2004034780-A.
PD	29-APR-2004.
XX	
PP	17-OCT-2002; 2002KR-00063379.
PR	17-OCT-2002; 2002KR-00063379.
XX	
PA	(BIOL-) BIOLEADERS CORP.
PA	(UTCH-) UNIT CHONSUN CO LTD.
PA	(KORE-) KOREA RES INST BIOSCIENCE & BIOTECHNOLOG.
PI	Boo HR, Ham GS, Hong SP, Jung CM, Kim CJ, Lee DG, Lee JS;
PI	Park YG, Sung MH;
XX	
XX	WPI; 2004-577380/56.
DR	
PT	Method for surface expression of peptides p5 and anal3 using pgs bca
PT	gene.
PS	
XX	Disclosure; SEQ ID NO 2; 25pp; Korean.
XX	
CC	The invention relates to a novel method for surface expression of
CC	peptides p5 and Anal3 using a poly-gamma-glutamic acid synthase (pgs) bca
CC	gene, thereby removing a purification process of peptides p5 and Anal3,
CC	and using lactic acid bacteria for the surface expression, so that
CC	peptide antibiotics can be cheaply and stably mass-produced. An
CC	expression vector pHECUB:pgsa-p5 comprises one or more genes encoding
CC	poly-gamma-glutamic acid synthase selected from pgsB, pgsC and pgsA, and
CC	a gene for dipolar peptide antibiotics having antimicrobial, antifungal
CC	and anticancer activities, wherein the dipolar peptide antibiotic has
CC	homology to the peptide p5 encoded by the nucleotide sequence set forth
CC	in ADV55735, or to the peptide Anal3 encoded by the nucleotide sequence
CC	set forth in ADV55737. The present sequence represents the B. subtilis
CC	poly-gamma-glutamic acid synthase C gene used in the invention.
XX	
SEQ	Sequence 447 BP; 116 A; 84 C; 94 G; 153 T; 0 U; 0 Other;
	Query Match 100.0%; Score 447; DB 13; Length 447;
	Best Local Similarity 100.0%; Pred. No. 2.7e-126;
	Matches 447; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
QY	1 ATGTTCCGATCGATTTCATTCATCCGACATTTTATAGTGTCATCTCGATTAAATTTT 60
DB	1 ATGTTCCGATCGATTTCATTCATCCGACATTTTATAGTGTCATCTCGATTAAATTTT 60
QY	61 GCGGAAAAAACAGGGGTCGTGCGCGGAGGACTTGTGTACCGGGATATTAGACTTGTG 120
DB	61 GCGGAAAAAACAGGGGTCGTGCGCGGAGGACTTGTGTACCGGGATATTAGACTTGTG 120
QY	121 TTAAATCAGCCGCTCTTATTTTACTTGTTTGTCTAGTAGCTTGCTCACTTAGTATC 180
DB	121 TTAAATCAGCCGCTCTTATTTTACTTGTTTGTCTAGTAGCTTGCTCACTTAGTATC 180
QY	181 GTGAATAATCGTTTATCCAAATTTATGATTTGTACGACGCAAAAAATTGCTGCATG 240
DB	181 GTGAATAATCGTTTATCCAAATTTATGATTTGTACGACGCAAAAAATTGCTGCATG 240
QY	241 CTGATTAACAGGGATCGTCTTAAAAATCGGGTTGATTTCTATACCCGATTTGACATT 300
DB	241 CTGATTAACAGGGATCGTCTTAAAAATCGGGTTGATTTCTATACCCGATTTGACATT 300
QY	301 GAAATCGCAGAAATTTGAGGAATCGGCATCATGTCACAGTTTAAATTCGAATACATT 360
DB	301 GAAATCGCAGAAATTTGAGGAATCGGCATCATGTCACAGTTTAAATTCGAATACATT 360
QY	361 CAGAAACAGGTTTAAACATTTACGTTTCGAGACGCTGTATTAGCGGAGGACCTTT 420
DB	361 CAGAAACAGGTTTAAACATTTACGTTTCGAGACGCTGTATTAGCGGAGGACCTTT 420

Db	361	CAGAACACAGCTTTAAACATTAGCTTCGGAAGACCGTGCTATTGAGCGGAGGACCTTT	420
QY	421	GCATCATGTTTGGTTACTACTTAATT	447
Db	421	GCATCATGTTTGGTTACTACTTAATT	447
RESULT 4			
ADA44796			
ID	ADA44796	standard; DNA; 6536 BP.	
XX	ADA44796;		
AC			
DT	20-NOV-2003	(first entry)	
XX			
DE	Microbial cell surface expression vector pGNBCA, SEQ ID NO:6.		
XX			
KW	Cell surface; expression vector; microbial;		
KW	poly-gamma-glutamate synthetase; Bacillus subtilis pgsBCA complex;		
KW	surface expression; Gram-positive bacterium; Gram-negative bacterium;		
KW	enzyme; antigen; antibody; attachment protein; adsorption protein;		
KW	vaccine; pGNBCA; cyclic; circular; ds.		
XX			
OS	Synthetic.		
XX	Bacillus subtilis; variety chungkookjang.		
XX			
XX	WO2003014360-A1.		
PD			
XX	20-FEB-2003.		
XX			
PF	09-AUG-2002; 2002WO-KR001522.		
XX			
PR	10-AUG-2001; 2001KR-00048373.		
XX			
PA	(BIOL-) BIOLEADERS CORP.		
XX	(MDMD-) MD LAB CO LTD.		
XX			
PI	Sung M, Hong S, Lee J, Jung C, Kim C, Soda K, Ashiuchi M;		
DR	WPI; 2003-256589/25.		
XX			
PT	New expression vector containing gene(s) that encode a poly-gamma-		
PT	glutamate synthetase complex, useful for producing proteins (e.g.		
PT	vaccines or enzymes) on the microbial surface of Gram-positive and/or		
PT	Gram-negative bacteria.		
XX			
PS	Example 1; Page 113-120; 122pp; English.		
XX			
CC	The invention relates to a vector for expression of a target protein on a		
CC	microbial cell surface. The vector of the invention comprises either one		
CC	or more than two genes of the Bacillus subtilis poly-gamma-glutamate		
CC	synthetase complex (pgsBCA) to facilitate microbial surface expression of		
CC	the target protein. The pgsBCA gene complex comprises the pgsB, pgsB and		
CC	pgsA genes and is normally expressed in the outer membrane of Bacillus		
CC	subtilis. The vector can be transformed into either Gram-positive or Gram		
CC	-negative bacteria (e.g., Escherichia coli), and can be used for the		
CC	surface expression of various proteins of interest such as enzymes,		
CC	antigens, antibodies, attachment proteins or adsorption proteins.		
CC	Proteins recombinantly produced using the vector of the invention can be		
CC	used as, for example, vaccines or enzymes. The present sequence		
CC	represents a microbial cell surface expression vector of the invention,		
CC	pGNBCA.		
XX			
XX			
SQ	Sequence 6536 BP; 1855 A; 1464 C; 1567 G; 1649 T; 0 U; 1 Other;		
QY	Query Match	100.0%; Score 447; DB 10; Length 6536;	
Db	Best Local Similarity	100.0%; Pred. No. 7,2e-126; Length 6536;	
	Matches 447; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
	1	ATGTCGATCAGATTATACATCGCAATAATTTAGTGTACTACAGTTTAATTTT	60
	1423	ATGTCGATCAGATTATACATCGCAATAATTTAGTGTACTACAGTTTAATTTT	1482

QY 61 GCGGAAAAACAGGATCGTCGCGCAGACGTTGTTGACCGGATATTAGACTTGTG 120
|
|
|
Db 1483 GCGGAAAAACAGGATCGTCGCGCAGACGTTGTTGACCGGATATTAGACTTGTG 1542
|
|
|
QY 121 TTTAATCAGCCGCTTTATTTACTTTGTTTGTGAGAGCTTGTCACTTAATGTTATC 180
|
|
|
Db 1543 TTTAATCAGCCGCTTTATTTACTTTGTTTGTGAGAGCTTGTCACTTAATGTTATC 1602
|
|
|
QY 181 GTGAATACGGTTTATCCAAATTTATGATTTGTACGAGCAGAAAAATTCGTCGCATG 240
|
|
|
Db 1603 GTGAATACGGTTTATCCAAATTTATGATTTGTACGAGCAGAAAAATTCGTCGCATG 1662
|
|
|
QY 241 CTGATTAACAGGATCGTCCTAAATAATCGCGTTGATTTTCTATACCCGATTTGATCAATT 300
|
|
|
Db 1663 CTGATTAACAGGATCGTCCTAAATAATCGCGTTGATTTTCTATACCCGATTTGATCAATT 1722
|
|
|
QY 301 GAAATCGCAGATTTCGAGAAATCGGCATCATCGTGCAGAGTTTATTTGCCAATTCATT 360
|
|
|
Db 1723 GAAATCGCAGATTTCGAGAAATCGGCATCATCGTGCAGAGTTTATTTGCCAATTCATT 1782
|
|
|
QY 361 CAGAAACAGAGTTTAAACATTACGTTGGAAAGCAGCGTCTATTGAGGCGAGCCATT 420
|
|
|
Db 1783 CAGAAACAGAGTTTAAACATTACGTTGGAAAGCAGCGTCTATTGAGGCGAGCCATT 1842
|
|
|
QY 421 GCTATCATGTTTGTCTACTTAATT 447
|
|
|
Db 1843 GCTATCATGTTTGTCTACTTAATT 1869
|
|
|

RESULT 5
AAF82256
ID AAF82256 standard; DNA; 450 BP.

AAF82256;

21-JUN-2001 (first entry)

Bacillus subtilis IFO 3336 DNA encoding a PGA synthesizing enzyme.

Bacterial; poly-gamma-L-glutamic acid; L-PGA;

KW poly-gamma-D,L-glutamic acid; PGA; glutamate racemase; ds.

XX Bacillus subtilis.

Key Location/Qualifiers

FT CDS 1..450

FT CDS /*tag= a

XX JP2001017182-A.

XX PD 23-JAN-2001.

XX PF 09-JUL-1999; 99JP-00196335.

XX PR 09-JUL-1999; 99JP-00196335.

XX PA (NACS) NAGASE SANGYO KK.

XX DR WPI; 2001-285408/30.

XX DR P-PSDB; AAB74025.

XX PT New nucleic acid encoding a glutamate racemase enzyme useful for the

XX PT preparation of poly-gamma-glutamic acid.

XX PS Claim 3; Page 13; 17pp; Japanese.

XX CC The present sequence encodes an enzyme which is useful in the production

XX CC of poly-gamma-L-glutamic acid (L-PGA) or poly-gamma-D,L-glutamic acid

XX CC (PGA). A plasmid comprising the present sequence may be used to transform

XX CC *Escherichia coli*. The transformants express the enzyme and PGA is

XX CC produced in the culture

XX CC Sequence 450 BP; 118 A; 84 C; 96 G; 152 T; 0 U; 0 Other;

Query Match 98.6%; Score 440.6; DB 4; Length 450;
Best Local Similarity 99.1%; Pred. No. 2,4e-124;
Matches 443; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATGTCGATCAGATTATATACATGCACTAATTTAGGTGATCACTCACTTAATTTT 60
|
|
|
Db 1 ATGTCGATCAGATTATATACATGCACTAATTTAGGTGATCACTCACTTAATTTT 60
|
|
|
QY 61 GCGGAAAAACAGGATCGTCGCGCAGACGTTGTTGACCGGATATTAGACTTGTG 120
|
|
|
Db 61 GCGGAAAAACAGGATCGTCGCGCAGACGTTGTTGACCGGATATTAGACTTGTG 120
|
|
|
QY 121 TTTAATCAGCCGCTTTATTTACTTTGTTTGTGAGAGCTTGTCACTTAATGTTATC 180
|
|
|
Db 121 TTTAATCAGCCGCTTTATTTACTTTGTTTGTGAGAGCTTGTCACTTAATGTTATC 180
|
|
|
QY 181 GTGAATACGGTTTATCCAAATTTATGATTTGTACGAGCAGAAAAATTCGTCGCATG 240
|
|
|
Db 181 GTGAATACGGTTTATCCAAATTTATGATTTGTACGAGCAGAAAAATTCGTCGCATG 240
|
|
|
QY 241 CTGATTAACAGGATCGTCCTAAATAATCGCGTTGATTTTCTATACCCGATTTGATCAATT 300
|
|
|
Db 241 CTGATTAACAGGATCGTCCTAAATAATCGCGTTGATTTTCTATACCCGATTTGATCAATT 300
|
|
|
QY 301 GAAATCGCAGATTTCGAGAAATCGGCATCATCGTGCAGAGTTTATTTGCCAATTCATT 360
|
|
|
Db 301 GAAATCGCAGATTTCGAGAAATCGGCATCATCGTGCAGAGTTTATTTGCCAATTCATT 360
|
|
|
QY 361 CAGAAACAGAGTTTAAACATTACGTTGGAAAGCAGCGTCTATTGAGGCGAGCCATT 420
|
|
|
Db 361 CAGAAACAGAGTTTAAACATTACGTTGGAAAGCAGCGTCTATTGAGGCGAGCCATT 420
|
|
|
QY 421 GCTATCATGTTTGTCTACTTAATT 447
|
|
|
Db 421 GCTATCATGTTTGTCTACTTAATT 447
|
|
|

RESULT 6
AAF82254
ID AAF82254 standard; DNA; 3045 BP.

AAF82254;

21-JUN-2001 (first entry)

Bacillus subtilis IFO 3336 DNA encoding glutamate racemase enzyme.

Bacterial; poly-gamma-L-glutamic acid; L-PGA;

KW poly-gamma-D,L-glutamic acid; PGA; glutamate racemase; ds.

XX Bacillus subtilis.

Key Location/Qualifiers

FT CDS 19..1200

FT CDS /*tag= a

FT CDS 1215..1264

FT CDS /*tag= b

FT CDS 1683..2825

FT CDS /*tag= c

XX JP2001017182-A.

XX PD 23-JAN-2001.

XX PF 09-JUL-1999; 99JP-00196335.

XX PR 09-JUL-1999; 99JP-00196335.

XX PA (NACS) NAGASE SANGYO KK.

XX DR WPI; 2001-285408/30.

XX DR P-PSDB; AAB74024, AAB74025, AAB74026.

XX New nucleic acid encoding a glutamate racemase enzyme useful for the
PT preparation of poly-gamma-glutamic acid.
XX
XX
PS Claim 1, Page 7-11, 17pp; Japanese.

XX The present sequence encodes a glutamate racemase enzyme which is useful
CC in the production of poly-gamma-L-glutamic acid (L-PGA) or poly-gamma-D,L
CC -glutamic acid (PGA). A plasmid comprising the present sequence may be
CC used to transform *Escherichia coli*. The transformants express the enzyme
CC and PGA is produced in the culture
XX

SO Sequence 3045 BP; 1000 A; 593 C; 696 G; 756 T; 0 U; 0 Other;

Query Match 98.6%; Score 440.6; DB 4; Length 3045;
Best Local Similarity 99.1%; Pred. No. 4,9e-124;
Matches 443; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATGTTCCGATCGATTATATACATCCGACATTAATTTAGTGTCTCTCACTTAATTTT 60
DB 1215 ATGTTCCGATCGATTATATACATCCGACATTAATTTAGTGTCTCTCACTTAATTTT 1274
QY 61 GCGGAAAAACAGGAGATCGTCCGCGAGACTTGTGTACCGGATATTAGACTTGTG 120
DB 1275 GCGGAAAAACAGGAGATCGTCCGCGAGACTTGTGTACCGGATATTAGACTTGTG 1334
QY 121 TTTAATCAGCCGCTTTTATTTTACTTGTGTGTCTAGTACGCTTGTCTCACTTAATTTT 180
DB 1335 TTTAATCAGCCGCTTTTATTTTACTTGTGTGTCTAGTACGCTTGTCTCACTTAATTTT 1394
QY 181 GTGAATACGCTTTATCCAAATTTATGATTTTGTATGAGACGCAAAATTTGTGCGCATG 240
DB 1395 GTGAATACGCTTTATCCAAATTTATGATTTTGTATGAGACGCAAAATTTGTGCGCATG 1454
QY 241 CTGATACAGGAGATCGTCTTAAATTCGCTTGTATTTCTATACCCGATTTGACATTT 300
DB 1455 CTGATACAGGAGATCGTCTTAAATTCGCTTGTATTTCTATACCCGATTTGACATTT 1514
QY 301 GAAATCGCAAAATTTGAGAAATCGGCATCATCTGCCAGTTTAAATTCGATATACAT 360
DB 1515 GAAATCGCAAAATTTGAGAAATCGGCATCATCTGCCAGTTTAAATTCGATATACAT 1574
QY 361 CAGAAACAAAGTTTAAACATTTACGTTTCGAAAGACGCTGCTATTTAGCGGAGCCTTT 420
DB 1575 CAGAAACAAAGTTTAAACATTTACGTTTCGAAAGACGCTGCTATTTAGCGGAGCCTTT 1634
QY 421 GCTATCATGTTTGTACTTAATTT 447
DB 1635 GCTATCATGTTTGTACTTAATTT 1661

RESULT 7
ABK75634
ID ABK75634 standard; DNA; 450 BP.

AC ABK75634;

DT 13-AUG-2002 (first entry)

XX *Bacillus licheniformis* genomic sequence tag (GST) #2925.

XX *Bacillus licheniformis* genomic sequenced tag; GST;

KM altered culture condition; environmental stress;

XX *Bacillus licheniformis*.

OS *Bacillus licheniformis*.

XX WO200229113-A2.

XX 11-Apr-2002.

XX 05-OCT-2001; 2001WO-US031437.

PR 06-OCT-2000; 2000US-00680598.

PR 27-MAR-2001; 2001US-0279526P.

XX (NOVO) NOVOZYMES BIOTECH INC.

PA (NOVO) NOVOZYMES AS.

XX Berka R, Clausen IG;
PI WPI; 2002-416684/44.

XX Monitoring differential expression of several genes in first *Bacillus*
PT cell relative to expression of same genes in one or more second *Bacillus*
PT cells, by using substrate containing *Bacillus* genomic sequenced tag
array.

XX Claim 4; SEQ ID NO 2925; 200pp; English.

CC The invention describes a method of monitoring differential expression of
CC genes in a first *Bacillus* cell relative to expression of the genes in
CC other *Bacillus* cells, comprising hybridising labelled nucleic acid probes
CC isolated from *Bacillus* cells to a substrate containing array of *Bacillus*
CC genomic sequenced tags (GST), examining the array, and determining
CC relative gene expression by an observed hybridisation reporter signal of
CC a spot in the array. The method is useful for measuring the expression of
CC genes in a first *Bacillus* cell relative to expression of the same genes
CC in one or more second *Bacillus* cells. The method is useful for monitoring
CC global expression of several genes from a *Bacillus* cell, discovering new
CC genes, identifying possible functions of unknown open reading frames and
CC monitoring gene copy number variation and stability. Monitoring changes
CC in expression of genes may be used to provide a representation of the way
CC in which *Bacillus* cells adapt to changes in culture conditions,
CC environmental stress or other physiological provocation. Extensive follow
CC up characterisation is unnecessary, when one spot on an array equals one
CC gene or one open reading frame, since sequence information is available.
CC This sequence represents a genomic sequence tag (GST) used in the method
CC of the invention. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at fep.wipo.int/pub/published_pat_sequences
XX

SO Sequence 450 BP; 112 A; 85 C; 102 G; 151 T; 0 U; 0 Other;

Query Match 64.7%; Score 289.2; DB 6; Length 450;
Best Local Similarity 78.0%; Pred. No. 5.1e-78;
Matches 348; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 1 ATGTTCCGATCGATTATATACATCCGACATTAATTTAGTGTCTCTCACTTAATTTT 60
DB 1 ATGTTCCGATCGATTATATATATCCGCTCATTTTAAAGAGTCTTCTCAGTTTGAATTTT 60
QY 61 GCGGAAAAACAGGAGATCGTCCGCGAGACTTGTGTACCGGATATTAGACTTGTG 120
DB 61 GCGGAAAAACAGGAGATCGTCCGCGAGACTTGTGTACCGGATATTAGACTTGTG 120
QY 121 TTTAATCAGCCGCTTTTATTTTACTTGTGTGTCTAGTACGCTTGTCTCACTTAATTTT 180
DB 121 TTTAATCAGCCGCTTTTATTTTACTTGTGTGTCTAGTACGCTTGTCTCACTTAATTTT 180
QY 181 GTGAATACGCTTTATCCAAATTTATGATTTTGTATGAGACGCAAAATTTGTGCGCATG 240
DB 181 GTGAATACGCTTTATCCAAATTTATGATTTTGTATGAGACGCAAAATTTGTGCGCATG 240
QY 241 CTGATACAGGAGATCGTCTTAAATTCGCTTGTATTTCTATACCCGATTTGACATTT 300
DB 241 CTGATACAGGAGATCGTCTTAAATTCGCTTGTATTTCTATACCCGATTTGACATTT 300
QY 301 GAAATCGCAAAATTTGAGAAATCGGCATCATCTGCCAGTTTAAATTCGATATACAT 360
DB 301 GAAATCGCAAAATTTGAGAAATCGGCATCATCTGCCAGTTTAAATTCGATATACAT 360
QY 361 CAGAAACAAAGTTTAAACATTTACGTTTCGAAAGACGCTGCTATTTAGCGGAGCCTTT 420
DB 361 CAGAAACAAAGTTTAAACATTTACGTTTCGAAAGACGCTGCTATTTAGCGGAGCCTTT 420

XX Sequence 459 BP; 135 A; 60 C; 82 G; 182 T; 0 U; 0 Other;

Query Match 42.4%; Score 189.4; DB 6; Length 459;

Best Local Similarity 64.0%; Pred. No. 1.8e-47;

Matches 286; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

```

OY 1 ATGTTGGATCAGATTATACATCGCACTAATTTAGTGTAAGTCTAGTTAATTTT 60
DB 7 ATGATAGGTTCAAGATTATATTTCTCTATTCGTAAGGTGTCGATCTCAGTTGATATT 66
OY 61 GCGGAAAAACAGGGATCGTCCCGAGAGACTGTTGTAACGGGATTTAGACCTTG 120
DB 67 GCTGGAATAATTTGGATTAAATCCACAGGGTAAAGTGTCCAGTTATTTAGCTTGATT 126
OY 121 TTATATCAGCCGCTTTATTTTACTGTTTGTAGTAGCTCTCACTTATGTTATC 180
DB 127 TTGATCAACCGATCATGTGTTATCAGATTAATCATTAAGTTGCTTAATTTATTC 186
OY 181 GTGAATAACGGTTTATCCAAATTTATGATTTTGTACGACGACGAAATTCCTGCATG 240
DB 187 GTAAACAACGGTATAGTAAGTGGTTATTTATGTAAGAAATAATTCCTGCATG 246
OY 241 CTGATTAACAGGATCGTCTTAATAATCGGCTTGAATTTCTATACCCGATTTACCATTT 300
DB 247 ATACTGACGGGAATGTAATTAATTTATTTATTTGATCTCTGATCCATTAACCCCATTT 306
OY 301 GAAATCGCAGAAATTCGAGAAATCGGCATCATGTCGAGTTTATTCGAATACCAT 360
DB 307 GAAATGTTGAAGTTTACAGTATAGGTGTGTCATTCCTGATTAATTCGAATACCAT 366
OY 361 CAGAAACAAGGTTTAAACATTAACGTTTCGAGACAGCTGCTATGAGCGGAGACCTTT 420
DB 367 CAAAAACAAGGTGATGATTAACATTTCTACAAACATGTTATTAACATGATTAACATAT 426
OY 421 GCTATCATGTTGTTTACTACTTAATT 447
DB 427 ATCATCTTATTTTATATAGTTTATT 453

```

RESULT 10

ADSO1974 standard; DNA; 459 BP.

ADSO1974;

04-NOV-2004 (first entry)

Staphylococcus epidermis polymnucleotide seqid 1269.

antibacterial; vaccine; antisense therapy; Staphylococcus epidermidis;
recombinant expression vector; infection; computer readable medium;
computer based system; gene; ds.

Staphylococcus epidermidis.

US2004147734-A1.

29-JUL-2004.

01-DEC-2003; 2003US-00724972.

08-NOV-1997; 97US-0064964P.

13-AUG-1998; 98US-00134001.

29-NOV-1999; 99US-00450969.

(DOUC/) DOUCETTE-STAMM L.

(BUSH/) BUSH D.

Doucette-Stamm L, Bush D;
WPI, 2004-580138/56.
P-Psdb; ADS05746.

XX New isolated polypeptide and encoding nucleic acid derived from
PT Staphylococcus epidermidis, useful for diagnosing, preventing and/or
PT treating an S. epidermidis bacterial infection.

PS Claim 5; SEQ ID NO 1269; 741bp; English.

XX The invention describes an isolated nucleic acid comprising a nucleotide
CC sequence with any of 3772 fully defined nucleotide sequences (SEQ ID NO:
CC 1-3772) and encoding an Staphylococcus epidermidis polypeptide with any
CC of 3772 fully defined amino acid sequences (SEQ ID NO: 3772-7544) as
CC a subject for S. epidermidis infection; a recombinant or substantially
CC vector; a cell comprising a recombinant expression vector of (1);
CC producing an S. epidermidis polypeptide; an isolated nucleic acid
CC comprising a nucleotide sequence of at least 8 nucleotides in length; a
CC vaccine composition for prevention or treatment of an S. epidermidis
CC infection, comprising a nucleic acid cited above and a carrier; treating
CC a subject for S. epidermidis infection; a recombinant or substantially
CC pure preparation of an S. epidermidis polypeptide or its fragment; a
CC vaccine composition for prevention or treatment of an S. epidermidis
CC infection; detecting the presence of a Staphylococcus nucleic acid in a
CC sample; a computer readable medium having recorded in it the nucleotide
CC sequences with SEQ ID NO: 1-3772 or its fragments; a computer based
CC system for identifying fragments of the Staphylococcus genome of
CC commercial importance; a computer based system for identifying fragments
CC of the Staphylococcus plasmidic acid fragments of the Staphylococcus
CC commercially important nucleic acid fragments of the Staphylococcus
CC genome and/or plasmids; and identifying an expression modulating fragment
CC of the Staphylococcus genome and/or plasmids. The methods and
CC compositions of the present invention are useful for the diagnosis,
CC prevention and/or treatment of an Staphylococcus epidermidis bacterial
CC infection. This sequence encodes a S. epidermis protein of the invention.

XX Sequence 459 BP; 135 A; 60 C; 82 G; 182 T; 0 U; 0 Other;
Query Match 42.4%; Score 189.4; DB 13; Length 459;
Best Local Similarity 64.0%; Pred. No. 1.8e-47;
Matches 286; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

```

OY 1 ATGTTGGATCAGATTATACATCGCACTAATTTAGTGTAAGTCTAGTTAATTTT 60
DB 7 ATGATAGGTTCAAGATTATATTTCTCTATTCGTAAGGTGTCGATCTCAGTTGATATT 66
OY 61 GCGGAAAAACAGGGATCGTCCCGAGAGACTGTTGTAACGGGATTTAGACCTTG 120
DB 67 GCTGGAATAATTTGGATTAAATCCACAGGGTAAAGTGTCCAGTTATTTAGCTTGATT 126
OY 121 TTATATCAGCCGCTTTATTTTACTGTTTGTAGTAGCTCTCACTTATGTTATC 180
DB 127 TTGATCAACCGATCATGTGTTATCAGATTAATCATTAAGTTGCTTAATTTATTC 186
OY 181 GTGAATAACGGTTTATCCAAATTTATGATTTTGTACGACGACGAAATTCCTGCATG 240
DB 187 GTAAACAACGGTATAGTAAGTGGTTATTTATATGTAAGAAATAATTCCTGCATG 246
OY 241 CTGATTAACAGGATCGTCTTAATAATCGGCTTGAATTTCTATACCCGATTTACCATTT 300
DB 247 ATACTGACGGGAATGTAATTAATTTATTTGATCTCTGATCCATTAACCCCATTT 306
OY 301 GAAATCGCAGAAATTCGAGAAATCGGCATCATGTCGAGTTTATTCGAATACCAT 360
DB 307 GAAATGTTGAAGTTTACAGTATAGGTGTGTCATTCCTGATTAATTCGAATACCAT 366
OY 361 CAGAAACAAGGTTTAAACATTAACGTTTCGAGACAGCTGCTATGAGCGGAGACCTTT 420
DB 367 CAAAAACAAGGTGATGATTAACATTTCTACAAACATGTTATTAACATGATTAACATAT 426
OY 421 GCTATCATGTTGTTTACTACTTAATT 447
DB 427 ATCATCTTATTTTATATAGTTTATT 453

```

RESULT 11

AAH54379
ID AAH54379 standard; DNA; 3240 BP.
XX
AC AAH54379;
XX
DT 03-SEP-2001 (first entry)
XX
DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:3743.
XX
KM Staphylococcus epidermidis SRI strain; infection; diagnosis; vaccination;
KM endocarditis; ds.
XX
OS Staphylococcus epidermidis.
XX
PN WO200134809-A2.
XX
XX 17-MAY-2001.
PD
XX 09-NOV-2000; 2000WO-US030782.
PF
XX 09-NOV-1999; 99US-0164258P.
PR
XX (GLAXO) GLAXO GROUP LTD.
PA
XX Kimmexly WJ;
PI
XX WPI; 2001-316495/33.
DR
XX Nucleic acid encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis.
PS
XX Claim 8; Page 1366-1367; 2188pp; English.
XX
XX AAH52304 to AAH53970 represent nucleic acid (I) encoding polypeptides
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis. (I)
CC and (II) can have antibacterial activity and therefore can be used in
CC vaccination. The nucleic acid (I) may be used to produce the S.
CC epidermidis polypeptides (II) via the production of vectors containing
CC them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acid) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to AAH55098
CC represent oligonucleotide sequences and primers which are used in the
CC exemplification of the present invention. N.B. The present invention
CC specifically claims all the polynucleotide sequences given in the
CC sequence listing of the present specification, however the sequence
CC listing only goes up to SEQ ID NO:4454 so even though sequences are given
CC in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present
CC for SEQ ID NO:4455 to 4464
XX
XX Sequence 3240 BP; 1128 A; 464 C; 589 G; 1059 T; 0 U; 0 Other;
SQ
Query March 40.7%; Score 182; DB 4; Length 3240;
Best Local Similarity 65.5%; Pred. No. 6.6e-45;
Matches 266; Conservative 0; Mismatches 140; Indels 0; Gaps 0;
QY 1 ATGTCGATCAGATTATATACATCCACTAATTTAGGTACTGATCAGTTTATTTT 60
DB 2834 ATGATAGTTGATGATTTATTTCTCTTATTCGATAGTGTGATCTCAGTTTGATTTT 2893
QY 61 GCGGAAAAACAGGATGTCGCGGAGACTGTTGTAACCGGATTTNAGACTGTG 120
DB 2894 GCTGGAATTTGGGATTAATCCACAGGTTAGCTTCCAGTTTATTTAGCTTTGATT 2953
QY 121 TTATATCGCGCGGTTTATTTTACTGTTTGTAGAGAGCTTCTCACTATCTTATC 180
DB 2954 TTGATCAACCGATATGTTGTTATCAGATTAATCAATTAGTCTTAACTTATTTATC 3013
QY 181 GTGAATACGGTTTATCCAAATTTATGATTTTGTACGAGCAGCAAAATTCGTCATG 240

DB 3014 GTAGCAACGGTATTAGTAAGTGGTATTATTTATGATGAGAAATTCGCGCCATG 3073
QY 241 CTGATTAACAGGATCCGCTTAAATAATCGCGTTGATTTTCTATACCGGATTTGACATT 300
DB 3074 ATACTGACGGGAATGGATTAATTAATTTATTTGATCTCTGTACCATTBACCCCATTT 3133
QY 301 GAATCCAGAAATTTCCAGGAATCGCATCATCGTCGAGGTTTAAATGGCAATACCAT 360
DB 3134 GAATGTTGTAAGTTTCAAGATATAGTGTTGTATCTCTGATATATTGGGAATACAA 3193
QY 361 CAGAAACAGGTTTAAACATTTAGCTTGGAAACACGCTGCTATTGA 406
DB 3194 CAAAACAGGTTGATCATTAACACTTTCTAACAAATGTTATTAA 3239
RESULT 12
ADB06111
ID ADB06111 standard; DNA; 480 BP.
XX
AC ADB06111;
XX
DT 20-NOV-2003 (first entry)
XX
DE Alloococcus otitis antigenic protein encoding DNA SEQ ID NO:51.
XX
KM Alloococcus otitis; antigenic protein; immunogenic; immunisation;
KM gene therapy; Gram-positive bacterium; infection; gene; ds.
XX
OS Alloococcus otitis.
XX
PN WO2003048304-A2.
XX
PD 12-JUN-2003.
XX
XX 25-NOV-2002; 2002WO-US036123.
PF
XX 29-NOV-2001; 2001US-0333777P.
PR 18-NOV-2002; 2002US-0426742P.
XX
PA (AMHP) WYETH HOLDINGS CORP.
PI
XX Fletcher LD, Mcmichael JC, Russell DP, Zagursky RJ;
XX WPI; 2003-505284/47.
DR P-PSDB; ADB06112.
XX
XX The present invention describes an isolated polynucleotide (I) of
CC Alloococcus otitis genomic DNA, which encodes an antigenic protein.
CC Alloococcus otitis is a Gram-positive bacterium. Also described: (1)
CC an isolated polypeptide that is encoded by the polynucleotide (I); (2) an
CC expression vector comprising the novel isolated polynucleotide (I), its
CC complement, degenerate variant or fragment; (3) a genetically engineered
CC host cell, transfected, transformed or infected with the vector of (2);
CC (4) an antibody specific for the polypeptide of (1); (5) an immunogenic
CC composition comprising the polypeptide, its complement, biological
CC equivalent or fragment, or the polynucleotide that is comprised in the
CC expression vector; (6) a pharmaceutical composition comprising the
CC polypeptide of (1) and a carrier; (7) a protein chip comprising an array
CC of the polypeptides of (1), their biological equivalent or fragment; (8)
CC immunising against Alloococcus otitis by administering to a host the
CC immunogenic composition; (9) detecting and/or identifying Alloococcus
CC otitis in the biological sample; (10) a kit comprising a container
CC containing the novel polynucleotide, its degenerate variant or fragment,
CC or the antibody of (4); and (11) producing a polypeptide by culturing the
CC genetically engineered host cell under conditions suitable to produce the
CC polypeptide from the culture. (I) can be used in gene therapy. The

CC polynucleotides, polypeptides, antibodies and compositions of the present
CC invention can be used for treating and diagnosing diseases, drug
CC screening assays and monitoring of effects during drug clinical trials.
CC The polynucleotides are useful for expressing and detecting *Alloisococcus*
CC *otitidis*. The present sequence encodes an *Alloisococcus otitidis* antigen
CC protein from the present invention.

XX Sequence 480 BP; 113 A; 78 C; 99 G; 190 T; 0 U; 0 Other;

Query Match 14.6%; Score 65.2; DB 9; Length 480;
Best Local Similarity 49.5%; Pred. No. 1.8e-09;
Matches 200; Conservative 0; Mismatches 196; Indels 6; Gaps 1;

```
QY 9 ATCAGATTATATACATGCGACTATTTAGTGTACTACTGACTTAAATTTTGGCGAAAA 68
DB 36 AACAGATTTTATATATGCTTGGCTTGGCTTAAACATATCCATTAATAGTGAAGCTT 95
QY 69 AACAGAGATTCGCGCGGAGCACTGTTGTACCGGGATTTTAAAGACTGTGTTAATCA 128
DB 96 TTTTGAATATATGCTGTGTATGATATGCCAGGATACCTAGCCCTTTATATGATGA 155
QY 129 GCCGCTTTATTTTACTGTTTGTGTAAGTGAAGCTTCACTTAATGTTATCGTGAATA 188
DB 156 GCCCTTGAATGCTTCTTCAGATATTTATCTACCTGCTAGCTAATTTGATTTAATTA 215
QY 189 CGGTTATTCAAATTTATGATTTTGTACGAGCAGAAAAATTCGCTGCATCTGATAC 248
DB 216 TGGGATTTGAAATTCGATTTTATATGGAAGCTTATTTGTTTAAATCTTGGT 275
QY 249 AGGATCGTCCCTTAAATGCGCTTGA-----TTTCTATACCGGATGTGACATTTGA 302
DB 276 AGCTTTGGCATTTAAATGATTAATTAAGTAACTATTTTATTTCTACTCTCTCCCTTGA 335
QY 303 AATCGAGATTTTCGAGATCGGCAATCATCTGCGAGGTTTAAATGCCAATACCATTTGA 362
DB 336 TTCGTTATGTTTAGGGGATTTGGGCAATGATTTCCCGTTTACTGCGCAATATCTATG 395
QY 363 GAAACAGAGTTTAAACCATTAAGCTTGGAAAGCAGCTGCTATTGA 406
DB 396 ACGACAGTCACTGCCGATTAACCTTTGGCTCGCTCATTTGGGATTTGA 439
```

RESULT 13
ADB06109
ID ADB06109 standard; DNA; 480 BP.

XX ADB06109;

XX 20-NOV-2003 (first entry)

DE *Alloisococcus otitidis* antigenic protein encoding DNA SEQ ID NO:49.

XX *Alloisococcus otitidis*; antigenic protein; immunogenic; immunisation;
KW gene therapy; Gram-positive bacterium; infection; gene; ds.

XX *Alloisococcus otitidis*.

XX MO2003048304-A2.

XX 12-JUN-2003.

XX 25-NOV-2002; 2002MO-US036123.

XX 29-NOV-2001; 2001US-033377P.

XX 18-NOV-2002; 2002US-0426742P.

XX (AMHP) WYETH HOLDINGS CORP.

XX Fletcher LD, McMichael JC, Russell DP, Zagursky RJ;

XX WPI; 2003-505284/47.

XX DR P-PSDB; ADB06110.

PT New *Alloisococcus otitidis* polynucleotides and polypeptides, useful for
PT treating and diagnosing diseases, drug screening assays and monitoring of
PT effects during drug clinical trials.

PS Claim 7; SEQ ID NO 49; 1019pp; English.

XX The present invention describes an isolated polynucleotide (1) of
CC *Alloisococcus otitidis* genomic DNA, which encodes an antigenic protein.
CC *Alloisococcus otitidis* is a Gram-positive bacterium. Also described: (1)
CC an isolated polypeptide that is encoded by the polynucleotide (1); (2) an
CC expression vector comprising the novel isolated polynucleotide (1), its
CC complement, degenerate variant or fragment; (3) a genetically engineered
CC host cell, transfected, transformed or infected with the vector of (2);
CC (4) an antibody specific for the polypeptide of (1); (5) an immunogenic
CC composition comprising the polypeptide, its complement, biological
CC equivalent or fragment, or the polynucleotide that is comprised in the
CC expression vector; (6) a pharmaceutical composition comprising the array
CC of polypeptides of (1) and a carrier; (7) a protein chip comprising an array
CC of the polypeptides of (1), their biological equivalent or fragment; (8)
CC immunising against *Alloisococcus otitidis* by administering to a host the
CC immunogenic composition; (9) detecting and/or identifying *Alloisococcus*
CC *otitidis* in the biological sample; (10) a kit comprising a container
CC containing the novel polynucleotide, its degenerate variant or fragment,
CC or the antibody of (4); and (11) producing a polypeptide by culturing the
CC genetically engineered host cell under conditions suitable to produce the
CC polypeptide from the culture. (1) can be used in gene therapy. The
CC polynucleotides, polypeptides, antibodies and compositions of the present
CC invention can be used for treating and diagnosing diseases, drug
CC screening assays and monitoring of effects during drug clinical trials.
CC The polynucleotides are useful for expressing and detecting *Alloisococcus*
CC *otitidis*. The present sequence encodes an *Alloisococcus otitidis* antigen
CC protein from the present invention.

XX Sequence 480 BP; 113 A; 78 C; 99 G; 190 T; 0 U; 0 Other;

Query Match 14.6%; Score 65.2; DB 9; Length 480;
Best Local Similarity 49.5%; Pred. No. 1.8e-09;
Matches 200; Conservative 0; Mismatches 196; Indels 6; Gaps 1;

```
QY 9 ATCAGATTATATACATGCGACTATTTAGTGTACTACTGACTTAAATTTTGGCGAAAA 68
DB 36 AACAGATTTTATATATGCTTGGCTTGGCTTAAACATATCCATTAATAGTGAAGCTT 95
QY 69 AACAGAGATTCGCGCGGAGCACTGTTGTACCGGGATTTTAAAGACTGTGTTAATCA 128
DB 96 TTTTGAATATATGCTGTGTATGATATGCCAGGATACCTAGCCCTTTATATGATGA 155
QY 129 GCCGCTTTATTTTACTGTTTGTGTAAGTGAAGCTTCACTTAATGTTATCGTGAATA 188
DB 156 GCCCTTGAATGCTTCTTCAGATATTTATCTACCTGCTAGCTAATTTGATTTAATTA 215
QY 189 CGGTTATTCAAATTTATGATTTTGTACGAGCAGAAAAATTCGCTGCATCTGATAC 248
DB 216 TGGGATTTGAAATTCGATTTTATATGGAAGCTTATTTGTTTAAATCTTGGT 275
QY 249 AGGATCGTCCCTTAAATGCGCTTGA-----TTTCTATACCGGATGTGACATTTGA 302
DB 276 AGCTTTGGCATTTAAATGATTAATTAAGTAACTATTTTATTTCTACTCTCTCCCTTGA 335
QY 303 AATCGAGATTTTCGAGATCGGCAATCATCTGCGAGGTTTAAATGCCAATACCATTTGA 362
DB 336 TTCGTTATGTTTAGGGGATTTGGGCAATGATTTCCCGTTTACTGCGCAATATCTATG 395
QY 363 GAAACAGAGTTTAAACCATTAAGCTTGGAAAGCAGCTGCTATTGA 406
DB 396 ACGACAGTCACTGCCGATTAACCTTTGGCTCGCTCATTTGGGATTTGA 439
```

RESULT 14
ADB12064_00/c
WP Sequence split into 18 fragments LOCUS ADB12064 Accession Adb12064
WP Fragment Name Begin End
WP ADB12064_00 1 110000

WP ADB12064_01 100001 210000
WP ADB12064_02 200001 310000
WP ADB12064_03 300001 410000
WP ADB12064_04 400001 510000
WP ADB12064_05 500001 610000
WP ADB12064_06 600001 710000
WP ADB12064_07 700001 810000
WP ADB12064_08 800001 910000
WP ADB12064_09 900001 1010000
WP ADB12064_10 1000001 1110000
WP ADB12064_11 1100001 1210000
WP ADB12064_12 1200001 1310000
WP ADB12064_13 1300001 1410000
WP ADB12064_14 1400001 1510000
WP ADB12064_15 1500001 1610000
WP ADB12064_16 1600001 1710000
WP ADB12064_17 1700001 1754382
ID ADB12064 standard; DNA; 1754382 BP.
XX
XX ADB12064;
AC
XX 20-NOV-2003 (first entry)
DT
XX
DE Alloiococcus ostitis entire genome sequence SEQ ID NO:6651.
XX
KM Alloiococcus ostitis; antigenic protein; immunogenic; immunisation;
KW gene therapy; Gram-positive bacterium; infection; gene; ds.
XX
OS Alloiococcus ostitis.
XX
PN WO2003048304-A2.
XX
PD 12-JUN-2003.
XX
XX 25-NOV-2002; 2002WO-US036123.
PF
XX 29-NOV-2001; 2001US-033777P.
PR 18-NOV-2002; 2002US-0426742P.
XX
PA (AMHP) WYETH HOLDINGS CORP.
XX
PI Fletcher JD, McMichael JC, Russell DP, Zagursky RJ;
XX
DR WPI; 2003-505284/47.
XX
XX New Alloiococcus ostitis polynucleotides and polypeptides, useful for
PT treating and diagnosing diseases, drug screening assays and monitoring of
PT effects during drug clinical trials.
XX
XX Example 3; SEQ ID NO 6651; 1019pp; English.
PS
XX The present invention describes an isolated polynucleotide (1) of
CC Alloiococcus ostitis genomic DNA, which encodes an antigenic protein.
CC Alloiococcus ostitis is a Gram-positive bacterium. Also described: (1)
CC an isolated polypeptide that is encoded by the polynucleotide (1); (2) an
CC expression vector comprising the novel isolated polynucleotide (1); its
CC complement, degenerate variant or fragment; (3) a genetically engineered
CC host cell, transfected, transformed or infected with the vector of (2);
CC (4) an antibody specific for the polypeptide of (1); (5) an immunogenic
CC composition comprising the polypeptide, its complement, biological
CC equivalent or fragment, or the polynucleotide that is comprised in the
CC expression vector; (6) a pharmaceutical composition comprising the
CC polypeptide of (1) and a carrier; (7) a protein chip comprising an array
CC of the polypeptides of (1), their biological equivalent or fragment; (8)
CC immunising against Alloiococcus ostitis by administering to a host the
CC immunogenic composition; (9) detecting and/or identifying Alloiococcus
CC ostitis in the biological sample; (10) a kit comprising a container
CC containing the novel polynucleotide, its degenerate variant or fragment,
CC or the antibody of (4); and (11) producing a polypeptide by culturing the
CC genetically engineered host cell under conditions suitable to produce the
CC polypeptide from the culture. (1) can be used in gene therapy. The
CC polynucleotides, polypeptides, antibodies and compositions of the present
-CC invention can be used for treating and diagnosing diseases, drug

CC screening assays and monitoring of effects during drug clinical trials.
CC The polynucleotides are useful for expressing and detecting Alloiococcus
CC ostitis. The present sequence represents the entire genome of
CC Alloiococcus ostitis, which is given in the exemplification of the
CC present invention.
XX
XX Sequence 1754382 BP; 484756A; 391265C; 387369G; 490992T; 0U; 0Other;
SQ
Query Match 14.6%; Score 65.2; DB 9; Length 110000;
Best Local Similarity 49.5%; Pred. No. 1.3e-08;
Matches 200; Conservative 0; Mismatches 198; Indels 6; Gaps 1;
OY 9 ATCAGATTATATACATCCACTAATTTAGGTGTAATCTCACTAGTTAAATTTTGGCGAAA 68
DB 22230 AACAGATTTTATATTCCTTGCTTGGCTTAACACTATCCCTATTAGTAGAAAGCTT 22171
OY 69 AACAGGATCGTCCCGCAGAGACTTTGTACCGGATATTAGAGACTTGTTTAATCA 128
DB 22170 TTTTGAATTATTCCTGGTGTATGATGCCCCAGATACCTACCTTATATTGATGA 22111
OY 129 GCCGCTCTTATTTACTTGTTTGTCTAGTAGAGTCTCACTATGTTATCGGAATA 188
DB 22110 GCCCTTGATGCTTCTTCAAGTATTTAATCTCACTGCTATGTTGATGTTTAATTA 22051
OY 189 CGGTTATCCAAATTTATGATTTTGTACGAGCAGAGAAATTCGCTGCATGCTGATPAC 248
DB 22050 TGGGATTTGAAATTCGTCATTTATATGGAACGTTATTTTAAATATCTTGGT 21991
OY 249 AGGATGCTCTTAAATCGCTTGA-----TTTCTATATCCGATGTGACATTGA 302
DB 21990 AGCTTGGCATTAATGATGATTAATGAACATATCTTATTCCTACTCTCCCTTAGA 21931
OY 303 AATCGCAGAAATTTGAGGAATCGGCATCATCGCCAGTTTATGTCATACATCA 362
DB 21930 TTCGTTATGTTTATGAGGAGATGCGGCAATGATTCGCTTACTGCGCAATCTATTCG 21871
OY 363 GAAACAGGTTTAAACATTACGTTGGAAGCAGCCTGCTATTTGA 406
DB 21870 ACGACAGTCACTGCCGATTAACCTTTGGCTGGCTCATTTGGGATTTGA 21827
RESULT 15
ABN92026/C
ID ABN92026 standard; DNA; 198 BP.
XX
XX ABN92026;
AC
XX 24-JUN-2002 (first entry)
DT
XX
DE Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:1469.
XX
XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
KW antibacterial; gene therapy; gene; ds.
XX
XX Staphylococcus epidermidis.
OS
XX US6380370-B1.
PN
XX 30-APR-2002.
PD
XX 13-AUG-1998; 98US-00134001.
PF
XX 14-AUG-1997; 97US-0055779P.
PR 08-NOV-1997; 97US-0064964P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
PA
XX Doucetie-Stamm LA, Bush D;
PI
XX WPI; 2002-381255/41.
DR P-PSDB; ABP39481.
XX
XX Novel isolated nucleic acid encoding a Staphylococcus epidermis

polypeptide, useful for diagnosing and treating bacterial infections.

Disclousure; SEQ ID NO 1489; 267bp; English.

ABN90538 to ABN93374 represent *Staphylococcus epidermidis* open reading frame (ORF) nucleic acid sequences which encode the amino acid sequences given in ABP3124 to ABP37960. The *S. epidermidis* sequences have antibacterial activity and can be used in gene therapy. The sequences can also be used in the diagnosis and treatment of bacterial infections, particularly *S. epidermidis* infections. The sequences can be used to screen for compounds able to interfere with the *S. epidermidis* life cycle or inhibit *S. epidermidis* infection. N.B. The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the USPTO web site

Sequence 198 BP; 61 A; 51 C; 29 G; 57 T; 0 U; 0 Other;

Query Match 11.8%; Score 52.8; DB 6; Length 198;

Best Local Similarity 69.2%; Pred. No. 8e-06; Mismatches 32; Indels 0; Gaps 0;

Matches 72; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```

Oy 1 ATGTCGATCAGATTATACATCGACTAATTTAGTGTACTACTCAGTTTAAATTTT 60
    |||||
Db 104 ATGATAGCTTCAGAAATTAATTTCTCTTATTCGTAGTGTGTCTCAGTTTGAATTT 45
    |||||
Oy 61 GCGAATAAACAAGGATCGTCCGCGCAGGACTTGTGTACCGGG 104
    |||||
Db 44 GCTGAGAAATTTGGATTAATCCAGCAGGTTAGTGTTCAGG 1
    |||||

```

Search completed: February 27, 2006, 07:05:18
Job time : 285.394 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:34:28 ; Search time 489.476 Seconds
(without alignments)
7551.777 Million cell updates/sec

Title: US-10-789-164-2

Perfect score: 447
Sequence: 1 atgttcgacatcagattata.....tgttttctactactaatt 447

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Published Applications NA.Main:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US03_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
6: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	447	100.0	447	US-10-484-605-2	Sequence 2, Appl1
2	447	100.0	447	US-10-789-164-2	Sequence 2, Appl1
3	447	100.0	6336	US-10-484-605-6	Sequence 6, Appl1
4	289.2	64.7	450	US-09-974-300-2925	Sequence 2925, Ap
5	189.4	42.4	459	US-10-724-972A-1269	Sequence 1269, Ap
6	65.2	14.6	480	US-10-501-282-49	Sequence 49, Appl
7	65.2	14.6	480	US-10-501-282-51	Sequence 51, Appl
8	65.2	14.6	1754382	US-10-501-282-6651	Sequence 6651, Ap
9	39.6	8.9	3673778	US-10-312-841-1	Sequence 1, Appl1
10	39.6	8.4	5822	US-10-311-455-1069	Sequence 1069, Ap
11	36.2	8.1	26997	US-10-221-714A-474	Sequence 474, Appl
12	36	8.1	7195	US-10-239-676-30	Sequence 30, Appl
13	36	8.1	7195	US-10-240-453-40	Sequence 40, Appl
14	35.8	8.0	358	US-10-425-115-69390	Sequence 69390, A
15	35.2	7.9	2100	US-09-974-300-221	Sequence 221, Appl
16	35.2	7.9	14078	US-09-928-847B-43	Sequence 43, Appl
17	35	7.8	1347	US-10-282-122A-37840	Sequence 37840, A
18	35	7.8	11839	US-09-764-891-7628	Sequence 7628, Ap
19	35	7.8	3673778	US-10-312-841-2	Sequence 2, Appl1
20	34.4	7.7	534	US-10-139-667-6	Sequence 6, Appl1
21	34.4	7.7	637	US-10-027-632-225214	Sequence 225214, A
22	34.4	7.7	637	US-10-027-632-225215	Sequence 225215, A
23	34.4	7.7	637	US-10-027-632-225214	Sequence 225214, A

24	34.4	7.7	637	US-10-027-632-225215	Sequence 225215, A
25	34.4	7.7	5769	US-10-257-166-101	Sequence 101, Appl
26	34.4	7.7	51259	US-10-374-077-209	Sequence 209, Appl
27	34.2	7.7	1344	US-10-472-928-2631	Sequence 2631, Appl
28	34.2	7.7	5273	US-10-311-455-847	Sequence 847, Appl
29	34.2	7.7	6052	US-10-311-455-190	Sequence 390, Appl
30	34.2	7.7	12666	US-08-961-527-137	Sequence 137, Appl
31	34.2	7.7	12666	US-10-158-844-137	Sequence 137, Appl
32	34.2	7.7	277616	US-10-367-094-83	Sequence 83, Appl1
33	34.2	7.7	216258	US-10-472-928-4979	Sequence 4979, Ap
34	34	7.6	1905	US-10-741-849-6083	Sequence 6083, Ap
35	34	7.6	14023	US-10-311-455-2078	Sequence 2078, Ap
36	34	7.6	34769	US-10-221-714A-502	Sequence 502, Appl
37	33.8	7.6	7046	US-10-311-455-2090	Sequence 2090, Ap
38	33.8	7.6	110079	US-10-175-523-96	Sequence 96, Appl1
39	33.8	7.6	110079	US-11-099-266-96	Sequence 96, Appl1
40	33.6	7.5	389	US-09-925-065A-307190	Sequence 307190, A
41	33.6	7.5	7173	US-11-097-143-22021	Sequence 22021, A
42	33.6	7.5	8303	US-10-311-455-395	Sequence 395, Appl
43	33.6	7.5	96109	US-10-672-787-35	Sequence 35, Appl
44	33.4	7.5	423	US-10-424-599-25187	Sequence 25187, A
45	33.4	7.5	460	US-09-918-995-26253	Sequence 26253, A

ALIGNMENTS

RESULT 1
US-10-484-605-2
Sequence 2, Application US/10484605
Publication No. US20040253704A1
GENERAL INFORMATION:
APPLICANT: Sung, Moon-Hee
APPLICANT: Hong, Seung-Pyo
APPLICANT: Lee, Jong-Su
APPLICANT: Jung, Chang-Min
APPLICANT: Kim, Chul-doong
APPLICANT: Soda, Kenji
APPLICANT: Ashuchi, Makoto
TITLE OF INVENTION: SURFACE EXPRESSION VECTORS HAVING pgsbca, THE GENE CODING
TITLE OF INVENTION: POLY-GAMMA-GLUTAMATE SYNTHETASE, AND A METHOD FOR EXPRESSION OF
TITLE OF INVENTION: TARGET PROTEIN AT THE SURFACE OF MICROORGANISM USING THE VECTOR
FILE REFERENCE: 4240-101
CURRENT FILING DATE: 2004-01-20
PRIOR FILING DATE: 2002-08-09
PRIOR APPLICATION NUMBER: PCT/KR02/01522
PRIOR FILING DATE: 2001-08-10
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn version 3.2
SEQ ID NO 2
LENGTH: 447
TYPE: DNA
ORGANISM: Bacillus subtilis
US-10-484-605-2

Query Match 100.0%; Score 447; DB 8; Length 447;
Best Local Similarity 100.0%; Pred. No. 1.6e-120;
Matches 447; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGTTGATCATGATTATTAACATCGACATTAATTTAGTGTAATCACTACTGATTAATTTT 60
DB 1 ATGTTGATCATGATTATTAACATCGACATTAATTTAGTGTAATCACTACTGATTAATTTT 60
QY 61 GCGGAAAAACGGGATCGCGCGGAGGACCTGTTGTAACCGGATATTAGGACTGTG 120
DB 61 GCGGAAAAACGGGATCGCGCGGAGGACCTGTTGTAACCGGATATTAGGACTGTG 120
QY 121 TTTAATCAGCGGCTTTTATTTACTGTTTGGTAGAGACTGCTGCTACTATGTTATC 180
DB 121 TTTAATCAGCGGCTTTTATTTACTGTTTGGTAGAGACTGCTGCTACTATGTTATC 180

181 GTGAATACGGTTATCCAAATTTATGATTTGTACGAGCCAGAAAATTCGTCGCATG 240
181 GTGAATACGGTTATCCAAATTTATGATTTGTACGAGCCAGAAAATTCGTCGCATG 240
241 CTGATACAGGATCGTCTTAAATTCGCGTTTGAATTTTCTATACCCGATTTGACATTT 300
241 CTGATACAGGATCGTCTTAAATTCGCGTTTGAATTTTCTATACCCGATTTGACATTT 300
301 GAAATCGCAGAAATTCGAGAAATCGGCATCATCGGCAAGGTTTAAATTCGCAATACCAT 360
301 GAAATCGCAGAAATTCGAGAAATCGGCATCATCGGCAAGGTTTAAATTCGCAATACCAT 360
361 CAGAAACAGGTTTAAACCATTAACGTTGCGAAGCAGCGTCTATATGAGCGAGCAGCTTT 420
361 CAGAAACAGGTTTAAACCATTAACGTTGCGAAGCAGCGTCTATATGAGCGAGCAGCTTT 420
421 GCTATCATGTTTGTACTACTTAAT 447
421 GCTATCATGTTTGTACTACTTAAT 447

RESULT 2

US-10-789-164-2
; Sequence 2, Application US/10789164
; Publication No. US20050191720A1
; GENERAL INFORMATION:
; APPLICANT: Sung, M.H. et al.
; TITLE OF INVENTION: SURFACE EXPRESSION METHOD OF PEPTIDES P5 AND ANAL3 USING THE GENE
; FILE REFERENCE: P1574
; CURRENT APPLICATION NUMBER: US/10/789,164
; CURRENT FILING DATE: 2004-02-27
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Knapentin 1.71
; SEQ ID NO 2
; LENGTH: 447
; TYPE: DNA
; ORGANISM: Bacillus subtilis
US-10-789-164-2

Query Match 100.0%; Score 447; DB 9; Length 447;
Best Local Similarity 100.0%; Pred. No. 1.6e-120;
Matches 447; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 ATGTTCCGATCAGATTTATACATCGCACTAATTTAGGATGATCTGATGATTTATTTT 60
1 ATGTTCCGATCAGATTTATACATCGCACTAATTTAGGATGATCTGATGATTTATTTT 60
61 GCGAAAAACAGGATCGTCCGCGCAGAGACTGTTGTACCGGATTTTGAAGACTGTG 120
61 GCGAAAAACAGGATCGTCCGCGCAGAGACTGTTGTACCGGATTTTGAAGACTGTG 120
121 TTTAATCAGCGGATCTTATTTTATTTTGTGATGAGTGTGCTCACTTATGATTC 180
121 TTTAATCAGCGGATCTTATTTTATTTTGTGATGAGTGTGCTCACTTATGATTC 180
181 GTGAATACGGTTATCCAAATTTATGATTTGTACGAGCCAGAAAATTCGTCGCATG 240
181 GTGAATACGGTTATCCAAATTTATGATTTGTACGAGCCAGAAAATTCGTCGCATG 240
241 CTGATACAGGATCGTCTTAAATTCGCGTTTGAATTTTCTATACCCGATTTGACATTT 300
241 CTGATACAGGATCGTCTTAAATTCGCGTTTGAATTTTCTATACCCGATTTGACATTT 300
301 GAAATCGCAGAAATTCGAGAAATCGGCATCATCGGCAAGGTTTAAATTCGCAATACCAT 360
301 GAAATCGCAGAAATTCGAGAAATCGGCATCATCGGCAAGGTTTAAATTCGCAATACCAT 360
361 CAGAAACAGGTTTAAACCATTAACGTTGCGAAGCAGCGTCTATATGAGCGAGCAGCTTT 420
361 CAGAAACAGGTTTAAACCATTAACGTTGCGAAGCAGCGTCTATATGAGCGAGCAGCTTT 420
421 GCTATCATGTTTGTACTACTTAAT 447

421 GCTATCATGTTTGTACTACTTAAT 447

RESULT 3

US-10-484-605-6
; Sequence 6, Application US/10484605
; Publication No. US20040253704A1
; GENERAL INFORMATION:
; APPLICANT: Sung, Moon-Hee
; APPLICANT: Hong, Seung-Pyo
; APPLICANT: Lee, Jong-Su
; APPLICANT: Jung, Chang-Min
; APPLICANT: Kim, Chul-Joong
; APPLICANT: Soda, Kenji
; APPLICANT: Ashiuchi, Makoto
; TITLE OF INVENTION: SURFACE EXPRESSION VECTORS HAVING pGBCA, THE GENE CODING
; TITLE OF INVENTION: POLY-GAMMA-GLUTAMATE SYNTHETASE, AND A METHOD FOR EXPRESSION OF
; FILE REFERENCE: 4240-101
; CURRENT APPLICATION NUMBER: US/10/484,605
; CURRENT FILING DATE: 2004-01-20
; PRIOR APPLICATION NUMBER: PCT/KR02/01522
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: KR 2001-48373
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 6
; LENGTH: 6536
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Construct
US-10-484-605-6

Query Match 100.0%; Score 447; DB 8; Length 6536;
Best Local Similarity 100.0%; Pred. No. 5.8e-120;
Matches 447; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 ATGTTCCGATCAGATTTATACATCGCACTAATTTAGGATGATCTGATGATTTATTTT 60
1423 ATGTTCCGATCAGATTTATACATCGCACTAATTTAGGATGATCTGATGATTTATTTT 1482
61 GCGAAAAACAGGATCGTCCGCGCAGAGACTGTTGTACCGGATTTTGAAGACTGTG 120
1483 GCGAAAAACAGGATCGTCCGCGCAGAGACTGTTGTACCGGATTTTGAAGACTGTG 1542
121 TTTAATCAGCGGATCTTATTTTATTTTGTGATGAGTGTGCTCACTTATGATTC 180
1543 TTTAATCAGCGGATCTTATTTTATTTTGTGATGAGTGTGCTCACTTATGATTC 1602
181 GTGAATACGGTTATCCAAATTTATGATTTGTACGAGCCAGAAAATTCGTCGCATG 240
1603 GTGAATACGGTTATCCAAATTTATGATTTGTACGAGCCAGAAAATTCGTCGCATG 1662
241 CTGATACAGGATCGTCTTAAATTCGCGTTTGAATTTTCTATACCCGATTTGACATTT 300
1663 CTGATACAGGATCGTCTTAAATTCGCGTTTGAATTTTCTATACCCGATTTGACATTT 1722
301 GAAATCGCAGAAATTCGAGAAATCGGCATCATCGGCAAGGTTTAAATTCGCAATACCAT 360
1723 GAAATCGCAGAAATTCGAGAAATCGGCATCATCGGCAAGGTTTAAATTCGCAATACCAT 1782
361 CAGAAACAGGTTTAAACCATTAACGTTGCGAAGCAGCGTCTATATGAGCGAGCAGCTTT 420
1783 CAGAAACAGGTTTAAACCATTAACGTTGCGAAGCAGCGTCTATATGAGCGAGCAGCTTT 1842
421 GCTATCATGTTTGTACTACTTAAT 447
1843 GCTATCATGTTTGTACTACTTAAT 1869

RESULT 4
US-09-974-300-2925
Sequence 2925, Application US/09974300
Patent No. US2002016721A1
GENERAL INFORMATION:
APPLICANT: Berka, Randy M.
TITLE OF INVENTION: Methods For Monitoring Multiple Gene
FILE REFERENCE: 10085,500-US
CURRENT APPLICATION NUMBER: US/09/974,300
PRIOR FILING DATE: 2001-10-05
PRIOR APPLICATION NUMBER: 09/680,598
PRIOR FILING DATE: 2000-10-06
PRIOR APPLICATION NUMBER: 60/279,526
PRIOR FILING DATE: 2001-03-27
NUMBER OF SEQ ID NOS: 8481
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2925
LENGTH: 450
TYPE: DNA
ORGANISM: Bacillus licheniformis
US-09-974-300-2925

Query Match 64.7%; Score 289.2; DB 3; Length 450;
Best Local Similarity 78.0%; Pred. No. 3.2e-74;
Matches 348; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 1 ATGTTGGATGAGTTTATACATGCACTAATTTAGGTGTAATTTAGTTAATTTT 60
DB 1 ATGTTGGATGAGTTTATATATATGCTCTATTTAGGAGTCTTACTGATTTT 60
QY 61 GCGAAAAACAGGAGATCGTCCGCGAGAGCTGTGTGTAACGGATATTTAGACTGTG 120
DB 61 GAGAGAAAAACGGGAATGTACAGCGCGCTGTGTGTAACGGATATTTAGACTGTG 120
QY 121 TTTAATCAGCGGTCTTATTTTACTGTTTGTCTAGTGTGCTGCTATGTAATC 180
DB 121 TTTAATCAGCGGTCTTATTTTACTGTTTGTCTAGTGTGCTGCTATGTAATC 180
QY 121 TTTAATCAGCGGTCTTATTTTACTGTTTGTCTAGTGTGCTGCTATGTAATC 180
DB 121 TTTAATCAGCGGTCTTATTTTACTGTTTGTCTAGTGTGCTGCTATGTAATC 180
QY 181 GTGAATAGGAGTTTATGCAATTTATGTTTGTGTAACGAGCAAAATTCGTCGATG 240
DB 181 GTGAATAGGAGTTTATGCAATTTATGTTTGTGTAACGAGCAAAATTCGTCGATG 240
QY 181 GTGAATAGGAGTTTATGCAATTTATGTTTGTGTAACGAGCAAAATTCGTCGATG 240
DB 181 GTGAATAGGAGTTTATGCAATTTATGTTTGTGTAACGAGCAAAATTCGTCGATG 240
QY 241 CTGATACAGGAGATCGTCTTAAATTCGCTTGTATTTCTATCCGATGATGCAATTT 300
DB 241 CTGATACAGGAGATCGTCTTAAATTCGCTTGTATTTCTATCCGATGATGCAATTT 300
QY 241 CTGATACAGGAGATCGTCTTAAATTCGCTTGTATTTCTATCCGATGATGCAATTT 300
DB 241 CTGATACAGGAGATCGTCTTAAATTCGCTTGTATTTCTATCCGATGATGCAATTT 300
QY 301 GAAATCGAGAAATTTGAGGAATCGGCAATCGTCCAGGTTTAAATTCGCAATTCATT 360
DB 301 GAAATCGAGAAATTTGAGGAATCGGCAATCGTCCAGGTTTAAATTCGCAATTCATT 360
QY 301 GAAATCGAGAAATTTGAGGAATCGGCAATCGTCCAGGTTTAAATTCGCAATTCATT 360
DB 301 GAAATCGAGAAATTTGAGGAATCGGCAATCGTCCAGGTTTAAATTCGCAATTCATT 360
QY 361 CAGAAACAGGTTTAACTTACGTTTCCGAGCAAGCTGATTTAGGAGGACCTTT 420
DB 361 CAGAAACAGGTTTAACTTACGTTTCCGAGCAAGCTGATTTAGGAGGACCTTT 420
QY 361 CAGAAACAGGTTTAACTTACGTTTCCGAGCAAGCTGATTTAGGAGGACCTTT 420
DB 361 CAGAAACAGGTTTAACTTACGTTTCCGAGCAAGCTGATTTAGGAGGACCTTT 420
QY 421 GCTATCATGTTTGTACTACTTAAT 446
DB 421 GCTATCATGTTTGTACTACTTAAT 446
QY 421 GCTATCATGTTTGTACTACTTAAT 446
DB 421 GCTATCATGTTTGTACTACTTAAT 446

RESULT 5
US-10-724-972A-1269
Sequence 1269, Application US/10724972A
Publication No. US2004014773A1
GENERAL INFORMATION:
APPLICANT: Doucette-Stamm, Lynn
APPLICANT: Bush, David
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: PTH03-16
CURRENT APPLICATION NUMBER: US/10/724,972A
CURRENT FILING DATE: 2003-12-01

PRIOR APPLICATION NUMBER: 09/450,969
PRIOR FILING DATE: 1999-11-29
PRIOR APPLICATION NUMBER: 09/134,001
PRIOR FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 7544
SEQ ID NO 1269
LENGTH: 459
TYPE: DNA
ORGANISM: S.epidermidis
US-10-724-972A-1269

Query Match 42.4%; Score 189.4; DB 7; Length 459;
Best Local Similarity 64.0%; Pred. No. 6.2e-45;
Matches 286; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

QY 1 ATGTTGGATGAGTTTATACATGCACTAATTTAGGTGTAATTTAGTTAATTTT 60
DB 7 ATGTTGGATGAGTTTATATATATGCTCTATTTAGGAGTCTTACTGATTTT 66
QY 61 GCGAAAAACAGGAGATCGTCCGCGAGAGCTGTGTGTAACGGATATTTAGACTGTG 120
DB 67 GCTAGAAAAATTTGGATTAATTCAGCAAGGTTAGTGTGTAACGGATATTTAGACTGTG 126
QY 121 TTTAATCAGCGGTCTTATTTTACTGTTTGTCTAGTGTGCTGCTATGTAATC 180
DB 127 TTTAATCAGCGGTCTTATTTTACTGTTTGTCTAGTGTGCTGCTATGTAATC 186
QY 181 GTGAATAGGAGTTTATGCAATTTATGTTTGTGTAACGAGCAAAATTCGTCGATG 240
DB 187 GTGAATAGGAGTTTATGCAATTTATGTTTGTGTAACGAGCAAAATTCGTCGATG 246
QY 241 CTGATACAGGAGATCGTCTTAAATTCGCTTGTATTTCTATCCGATGATGCAATTT 300
DB 247 ATGATGAGGAGATGATTAATTTATGTTTGTGTAACGAGCAAAATTCGTCGATG 306
QY 301 GAAATCGAGAAATTTGAGGAATCGGCAATCGTCCAGGTTTAAATTCGCAATTCATT 360
DB 307 GAAATCGAGAAATTTGAGGAATCGGCAATCGTCCAGGTTTAAATTCGCAATTCATT 366
QY 361 CAGAAACAGGTTTAACTTACGTTTCCGAGCAAGCTGATTTAGGAGGACCTTT 420
DB 367 CAGAAACAGGTTTAACTTACGTTTCCGAGCAAGCTGATTTAGGAGGACCTTT 426
QY 421 GCTATCATGTTTGTACTACTTAAT 447
DB 427 ATGATGATTTTATATATGTTTATTT 453

RESULT 6
US-10-501-282-49
Sequence 49, Application US/10501282
Publication No. US20050203280A1
GENERAL INFORMATION:
APPLICANT: MICHAEL, JOHN CALHOUN
APPLICANT: ZAGURSKY, ROBERT JOHN
APPLICANT: RUSSELL, DAVID PARRISH
APPLICANT: FLETCHER, LEAH DIANE
TITLE OF INVENTION: ALLOTOCOCUS OTITIDIS OPEN READING FRAMES (ORFs) ENCODING
TITLE OF INVENTION: POLYPEPTIDE ANTIGENS, IMMUNOGENIC COMPOSITIONS AND USES THEREOF
FILE REFERENCE: AML00780 L2
CURRENT APPLICATION NUMBER: US/10/501,282
PRIOR FILING DATE: 2004-07-09
PRIOR APPLICATION NUMBER: 60/333,777
PRIOR FILING DATE: 2001-11-29
PRIOR APPLICATION NUMBER: 60/426,742
PRIOR FILING DATE: 2002-11-18
PRIOR APPLICATION NUMBER: PCT/US02/36123
PRIOR FILING DATE: 2002-11-25
NUMBER OF SEQ ID NOS: 6653

DB 129 GCCGCTTTATTTACTTTGTTTCTTCTAGTACGCTTCTCACTTATGTTATCGTAATA 188
DB 22110 GCCCTTGATGCTTCTCAGTATATTTATCTCACTGCTAGTATTTGATTTGTTAATTA 22051
QY 189 CGGTTTATCCAAATTTATGATTTTGTACGAGCAGCAAAATTCGCTGCATCTGATAAC 248
DB 22050 TGGGATTCGAAATTCGATTTATATGAGGAAGCGTTATTTGTTTAAATATCTTGGT 21991
QY 249 AGGATCGTCTTAAATTCGCTTTGA-----TTTCTATACCCGATTTGACATTTGA 302
DB 21990 AGCTTGGCATTAATATGATTTATTAAGTAATTTCTATTTCTACTCTCCCTTCCCTAGA 21931
QY 303 AATGCGAATTTGAGAGATCGCATCATCTGCGAGTTTAAATTTGCCAATACCATTA 362
DB 21930 TTCGGTTATGTTTAGGGGATTTGGGCGAGTAGTCTCCGGTTTACGCGCAATATCTATCG 21871
QY 363 GAAACAGGTTTAACTTACGTTGGAAGCAAGCGCTGCTATTGA 406
DB 21870 ACGACATCATCTCCGATTAATTTGCTGCTGCTATTTGGGATTGA 21827

RESULT 9
US-10-312-841-1
; Sequence 1, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
; FILE REFERENCE: E01/1208/NO
; CURRENT APPLICATION NUMBER: US/10/312, 841
; PRIORITY FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 1
; LENGTH: 3673778
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; NAME/KEY: unsure
; LOCATION: (3294164)
; US-10-312-841-1

Query Match 8.9%; Score 39.6; DB 6; Length 3673778;
Best Local Similarity 44.0%; Pred. No. 38;
Matches 168; Conservative 0; Mismatches 214; Indels 0; Gaps 0;

QY 1 ATGTTGATCAGATTATTAATCACTGACACTTAATTTAGTGTACTCACTGTTAATTTT 60
DB 1316265 ATTAATAGTTTGAATTAATTAAGTTTATTTAGTTTAAATGATTAATAGTTT 1316324
QY 61 GCGGAAAAAAGGAGTCGCGCGGAGGACTGTTGTACCGGGAATTTAGACTTGTG 120
DB 1316325 ATGAAATTTTGAATTAATTAATGTTTATTAATGTTTGAATTAATTAATGTTT 1316384
QY 121 TTTAATCAGCGGCTTTATTTACTGTTTCTAGTAGTGTGCTCACTTAATGTTATC 180
DB 1316385 ATGTAGGTTTGAATTTATTTAGTTTATAGTAGTGTGAGATTAATTAATTAAT 1316444
QY 181 GTGAAATACGTTTATTCAAATTTATGATTTTGTACGAGCAGCAAAATTCGCTCCAG 240
DB 1316445 TTTAATGAGGTTTGAATTAATTAATGTTTATTAATGAGTTTGTGATTAATTAAT 1316504
QY 241 CTGATACAGAGATCGCTTAAATTCGCGTTGATTTCTAATACCGGATTTGACATTT 300
DB 1316505 TTTATGAAAGTTTGAATTAATTAATGTTTATTAATGAGTTTGTGATTAATTAAT 1316564
QY 301 GAATCGCAGAAATTTGAGAGATCGCATCATCTGCGAGTTTAAATTTGCCAATACCAT 360
DB 1316565 TTTATGAGGTTTGAATTAATTAATTAATTAATTAATGAGTTTGTGATTAATTAAT 1316624
QY 361 CAGAAACAGGTTTAACTATTA 382

DB 1316625 TTTATGAGGTTTGAATTA 1316646

RESULT 10
US-10-311-455-1069
; Sequence 1069, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determini
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311, 455
; PRIORITY FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1069
; LENGTH: 5822
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; US-10-311-455-1069

Query Match 8.4%; Score 37.4; DB 6; Length 5822;
Best Local Similarity 57.1%; Pred. No. 8.1;
Matches 68; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 106 TATTAGAGCTGTGTTAATCAGCGGCTTATTTACTGTTTGTGCTAGTACGCTTG 165
DB 4775 TATGTGAAGTTGTTTAAATATAGTTTGTGTTATTTGTTAGGTATTTTGT 4834
QY 166 CTCACCTATGTTATCGTGAATATACGTTTATTCAAATTTATGATTTTGTACGAGCGAG 224
DB 4835 TTTATTTAGTTTATGAAATTTTATTAATTTATTAATTTTATTAAGTTTATTAAGTAG 4893

RESULT 11
US-10-221-714A-474
; Sequence 474, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; TITLE OF INVENTION: Diagnosis of Diseases Associated with
; FILE REFERENCE: 5013.1005
; CURRENT APPLICATION NUMBER: US/10/221, 714A
; PRIORITY FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: PCT/EP01/02955
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: DE 10013847.0
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIORITY FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 474
; LENGTH: 26997

```

; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically created genomic DNA (Homo sapiens)
US-10-221-714A-474
Query Match      8.1%; Score 36.2; DB 7; Length 26997;
Best Local Similarity 52.3%; Pred. No. 37;
Matches      80; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY      67 AAACAGGAGATGTCGCGCAGGACTTGTGTACCGGATTTAGACCTGTGTTAAT 126
DB      15921 AAGAAAGGAGAGCGGTGTGTGATTCGAAAGTTGAAATGATGTTTGTGTTATA 15980

QY      127 CAGCGGTCTTATTTTACTGTTTGTAGAGACTTGCCTATGTTATCGTGAA 186
DB      15981 TAACGTGAATTAAGATTAATGTTTGTGAAGAAATGTTGTTGATTTGATGTTAAG 16040

QY      187 TACGCTTATCCAAATTTATGATTTTGTACGGA 219
DB      16041 TTAATTTATAAGATATTTTAACTTGTTGCTTA 16073

RESULT 12
US-10-239-676-30
; Sequence 30, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 5013.1003
; CURRENT FILING DATE: 2002-09-24
; PRIOR FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/EP01/03968
; DE 10019058.8
; DE 10032529.7
; DE 1003826.1
; DE 10043826.1
; PRIOR FILING DATE: 2001-04-06
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 30
; LENGTH: 7195
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically created genomic DNA (Homo sapiens)
US-10-239-676-30
Query Match      8.1%; Score 36; DB 5; Length 7195;
Best Local Similarity 54.5%; Pred. No. 23;
Matches      72; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY      84 GGCAGGACTGTGTACCGGATTTAGAGACTGTGTTAATCAGCGGCTTTATTT 143
DB      4814 GCGCGAGATTATCGATTTTGTGTTTATGCTTTTATTTATGATTAATGATTTATTT 4873

QY      144 ACTTTTTCAGAGAGCTTCTCACTATGTTATCGGAATACGTTTATCCAAAT 203
DB      4874 ATTAGTTATGTTGATTAATTTAATTAATTAAGTTAAGTTAATTTTGAATA 4933

QY      204 TATGATTTGTA 215
DB      4934 TGTAACTTAGTA 4945

RESULT 13
US-10-240-453-40
```

```

; Sequence 40, Application US/10240453
; Publication No. US20030148326A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA
; TITLE OF INVENTION: Transcription
; TITLE OF INVENTION: by Means of Assessing the Methylation Status of Genes Associated
; FILE REFERENCE: 5013.1009
; CURRENT FILING DATE: US/10/240.453
; PRIOR FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03973
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 350
; SEQ ID NO 40
; LENGTH: 7195
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically created genomic DNA (Homo sapiens)
US-10-240-453-40
Query Match      8.1%; Score 36; DB 6; Length 7195;
Best Local Similarity 54.5%; Pred. No. 23;
Matches      72; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY      84 GGCAGGACTGTGTACCGGATTTAGAGACTGTGTTAATCAGCGGCTTTATTT 143
DB      4814 GCGCGAGATTATCGATTTTGTGTTTATGCTTTTATTTATGATTAATGATTTATTT 4873

QY      144 ACTTTTTCAGAGAGCTTCTCACTATGTTATGTTAATACGTTTATCCAAAT 203
DB      4874 ATTAGTTATGTTGATTAATTTAATTAATTAAGTTAAGTTAATTTTGAATA 4933

QY      204 TATGATTTGTA 215
DB      4934 TGTAACTTAGTA 4945

RESULT 14
US-10-425-115-69390/c
; Sequence 69390, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21 (53222)B
; CURRENT FILING DATE: US/10/425.115
; PRIOR FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 69390
; LENGTH: 358
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MNT4577_163278C.1
US-10-425-115-69390
Query Match      8.0%; Score 35.8; DB 8; Length 358;
```

Best Local Similarity 56.3%; Pred. No. 6.4;
Matches 67; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 48 CAGTTTAATTTTGGGAAAAAAGAGGATCGTCCGAGACTTGTGACGGGATA 107
DB 166 CAGCTTTTTCAGTGGGATTAAGTGAACCGGGGGGGGCTTTCCCTGGGGAGG 107
QY 108 TTTAGACTTGTGTTAATCAGCCGCTTTAATTTTACTGTTTGTAGAGCTTGC 166
DB 106 GTTATTAACACCTGTAAATGGGGCGTTTCTTTTCTTTTGAAGATTAACTGGC 48

RESULT 15

US-09-974-300-221
; Sequence 221, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods for Monitoring Multiple Gene
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIOR FILING DATE: 2001-10-05
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 221
; LENGTH: 2100
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-221

Query Match 7.9%; Score 35.2; DB 3; Length 2100;
Best Local Similarity 52.8%; Pred. No. 22;
Matches 76; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 203 TTATGATTTTGTACGAGCAGAAAATTCGCTGCATGATTAACAGGATGCTCTTA 262
DB 1492 TTACCAATTTTAACTGTCTTTGCCCCATATGACTGCCGAGAAAAGATGCTTCC 1551
QY 263 AATCGCGTTTGAATTTCTATACCGATTTGATTTGAAATCGAGAAATTTGAGGAA 322
DB 1552 AATCGCGTGGCGCTCCGGAAGAGTTGAAGCGCTTGAAGAAAAGCAGGAATTCATCA 1611
QY 323 TCGGCATCATCGTCCAGGTTTA 346
DB 1612 TTCAGGTCGATGAACCTGCCTTGA 1635

Search completed: February 27, 2006, 07:58:32.
Job time : 498.476 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:37:00 ; Search time 269.374 Seconds
(without alignments)
3538.457 Million cell updates/sec

Title: US-10-789-164-2

Perfect score: 447
Sequence: 1 atgttcgacatcagattata.....tgttcttactactaatt 447

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 7209121 seqs, 1066183437 residues

Total number of hits satisfying chosen parameters: 14418242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA New:*
1: /cgn2_6/prodata/1/pubpna/US08_NEW_PUB.seq:*
2: /cgn2_6/prodata/1/pubpna/US06_NEW_PUB.seq:*
3: /cgn2_6/prodata/1/pubpna/US07_NEW_PUB.seq:*
4: /cgn2_6/prodata/1/pubpna/PCT_NEW_PUB.seq:*
5: /cgn2_6/prodata/1/pubpna/US09_NEW_PUB.seq1:*
6: /cgn2_6/prodata/1/pubpna/US09_NEW_PUB.seq1:*
7: /cgn2_6/prodata/1/pubpna/US10_NEW_PUB.seq1:*
8: /cgn2_6/prodata/1/pubpna/US10_NEW_PUB.seq1:*
9: /cgn2_6/prodata/1/pubpna/US11_NEW_PUB.seq2:*
10: /cgn2_6/prodata/1/pubpna/US11_NEW_PUB.seq3:*
11: /cgn2_6/prodata/1/pubpna/US11_NEW_PUB.seq3:*
12: /cgn2_6/prodata/1/pubpna/US11_NEW_PUB.seq4:*
13: /cgn2_6/prodata/1/pubpna/US60_NEW_PUB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	44.7	100.0	447	US-10-530-083-2	Sequence 2, Appl1
2	189.4	42.4	453	US-10-793-626-2323	Sequence 2323, Ap
3	182	40.7	3240	US-10-793-626-3743	Sequence 3743, Ap
4	45.6	10.2	3029	US-10-793-626-4197	Sequence 4197, Ap
5	37.8	8.5	168753	US-11-181-234-1	Sequence 1, Appl1
6	33.6	7.5	498	US-09-925-065A-107190	Sequence 307190,
7	33.4	7.5	498	US-09-925-065A-297021	Sequence 297021,
8	33.4	7.5	528	US-09-925-065A-780420	Sequence 780420,
9	33.4	7.5	530	US-09-925-065A-804269	Sequence 804269,
10	33.4	7.5	569	US-09-925-065A-547674	Sequence 547674,
11	33.4	7.5	569	US-09-925-065A-547675	Sequence 547675,
12	33.4	7.5	611	US-09-925-065A-418523	Sequence 418523,
13	33.4	7.5	611	US-09-925-065A-418524	Sequence 418524,
14	33.4	7.5	1524	US-09-925-065A-692556	Sequence 692556,
15	33.4	7.5	1524	US-09-925-065A-692557	Sequence 692557,
16	33.4	7.5	1524	US-09-925-065A-692558	Sequence 692558,
17	33.4	7.5	1524	US-09-925-065A-692559	Sequence 692559,
18	33.4	7.5	1524	US-09-925-065A-692560	Sequence 692560,
19	33.4	7.5	1524	US-09-925-065A-692561	Sequence 692561,
20	33.4	7.5	2336	US-09-925-065A-674512	Sequence 674512,

ALIGNMENTS

RESULT 1	US-10-530-083-2	Sequence 2, Appl1	US/10530083	US-10-530-083-2	Sequence 2, Appl1
1	100.0%	Score 447;	DB 8;	Length 447;	Sequence 674513,
2	100.0%	Score 447;	DB 8;	Length 447;	Sequence 709850,
3	100.0%	Score 447;	DB 8;	Length 447;	Sequence 709851,
4	100.0%	Score 447;	DB 8;	Length 447;	Sequence 709852,
5	100.0%	Score 447;	DB 8;	Length 447;	Sequence 709853,
6	100.0%	Score 447;	DB 8;	Length 447;	Sequence 709854,
7	100.0%	Score 447;	DB 8;	Length 447;	Sequence 709855,
8	100.0%	Score 447;	DB 8;	Length 447;	Sequence 709856,
9	100.0%	Score 447;	DB 8;	Length 447;	Sequence 709857,
10	100.0%	Score 447;	DB 8;	Length 447;	Sequence 709858,
11	100.0%	Score 447;	DB 8;	Length 447;	Sequence 709859,
12	100.0%	Score 447;	DB 8;	Length 447;	Sequence 709860,
13	100.0%	Score 447;	DB 8;	Length 447;	Sequence 709861,
14	100.0%	Score 447;	DB 8;	Length 447;	Sequence 709862,
15	100.0%	Score 447;	DB 8;	Length 447;	Sequence 709863,
16	100.0%	Score 447;	DB 8;	Length 447;	Sequence 709864,
17	100.0%	Score 447;	DB 8;	Length 447;	Sequence 709865,
18	100.0%	Score 447;	DB 8;	Length 447;	Sequence 709866,
19	100.0%	Score 447;	DB 8;	Length 447;	Sequence 709867,
20	100.0%	Score 447;	DB 8;	Length 447;	Sequence 709868,
21	100.0%	Score 447;	DB 8;	Length 447;	Sequence 709869,
22	100.0%	Score 447;	DB 8;	Length 447;	Sequence 709870,
23	100.0%	Score 447;	DB 8;	Length 447;	Sequence 709871,
24	100.0%	Score 447;	DB 8;	Length 447;	Sequence 709872,
25	100.0%	Score 447;	DB 8;	Length 447;	Sequence 709873,
26	100.0%	Score 447;	DB 8;	Length 447;	Sequence 709874,
27	100.0%	Score 447;	DB 8;	Length 447;	Sequence 709875,
28	100.0%	Score 447;	DB 8;	Length 447;	Sequence 709876,
29	100.0%	Score 447;	DB 8;	Length 447;	Sequence 709877,
30	100.0%	Score 447;	DB 8;	Length 447;	Sequence 709878,
31	100.0%	Score 447;	DB 8;	Length 447;	Sequence 709879,
32	100.0%	Score 447;	DB 8;	Length 447;	Sequence 709880,
33	100.0%	Score 447;	DB 8;	Length 447;	Sequence 709881,
34	100.0%	Score 447;	DB 8;	Length 447;	Sequence 709882,
35	100.0%	Score 447;	DB 8;	Length 447;	Sequence 709883,
36	100.0%	Score 447;	DB 8;	Length 447;	Sequence 709884,
37	100.0%	Score 447;	DB 8;	Length 447;	Sequence 709885,
38	100.0%	Score 447;	DB 8;	Length 447;	Sequence 709886,
39	100.0%	Score 447;	DB 8;	Length 447;	Sequence 709887,
40	100.0%	Score 447;	DB 8;	Length 447;	Sequence 709888,
41	100.0%	Score 447;	DB 8;	Length 447;	Sequence 709889,
42	100.0%	Score 447;	DB 8;	Length 447;	Sequence 709890,
43	100.0%	Score 447;	DB 8;	Length 447;	Sequence 709891,
44	100.0%	Score 447;	DB 8;	Length 447;	Sequence 709892,
45	100.0%	Score 447;	DB 8;	Length 447;	Sequence 709893,

```
OY 181 GTGAAATACGGTTTATCCAAATTTATGATTTTGTACGAGCAGAGAAATTCGCTCCANG 240
DB 181 GTGAAATACGGTTTATCCAAATTTATGATTTTGTACGAGCAGAGAAATTCGCTCCATG 240
OY 241 CTGATTAACAGGAGATCGCTTAAATAATCGCGTTTGAATTTTCTATACCGGATTTGATACATT 300
DB 241 CTGATTAACAGGAGATCGCTTAAATAATCGCGTTTGAATTTTCTATACCGGATTTGATACATT 300
OY 301 GAAATCGAGAAATTTTCGAGAAATTCGGCATTCGTCGCAAGTTTAATTCGCAATCCATT 360
DB 301 GAAATCGAGAAATTTTCGAGAAATTCGGCATTCGTCGCAAGTTTAATTCGCAATCCATT 360
OY 361 CAGAAACAGGTTTAAACATTACGTTTCGAGAGACGCTGCTATTGAGCGAGCAGCTTT 420
DB 361 CAGAAACAGGTTTAAACATTACGTTTCGAGAGACGCTGCTATTGAGCGAGCAGCTTT 420
OY 421 GCTATCATGTTTGTCTTACTACTTAATT 447
DB 421 GCTATCATGTTTGTCTTACTACTTAATT 447
```

```
RESULT 2
US-10-793-626-2323
; Sequence 2323, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMBERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2323
; LENGTH: 453
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-10-793-626-2323
```

```
Query Match 42.4%; Score 189.4; DB 8; Length 453;
Best Local Similarity 64.0%; Pred. No. 4.5e-43;
Matches 286; Conservative 0; Mismatches 161; Indels 0; Gaps 0;
```

```
OY 1 ATGTTCCGATCAGATTTTATCATCGACCTAATTTTAAAGTGTACTCTCACTTAATTTT 60
DB 1 ATGATAGGTTCCAGAAATTAATTTCTCTTATTCGTAAGGTGCTGACTCAATTTGATTTT 60
OY 61 GCGGAAAAACAGGAGATCGTCCGAGAGACTTGTGTACCGGATTTTAACTGTG 120
DB 61 GCTGAGAAATTTGGATTAATCCACAGGTTAAGTGTTCACAGTTATTTAGCTTTGATT 120
OY 121 TTTAATCAGCGGCTTTATTTTACTTGTGTGTCTAGTAGAGCTTCTCACTTAATTC 180
DB 121 TTTGATCAACCGATCATGTTGTATTCAGATTAATCAATTAAGTCTTAATTTTATC 180
OY 181 GTGAAATACGGTTTATCCAAATTTATGATTTTGTACGAGCAGAGAAATTCGCTCCANG 240
DB 181 GTGAAATACGGTTTATCCAAATTTATGATTTTGTACGAGCAGAGAAATTCGCTCCANG 240
OY 241 CTGATTAACAGGAGATCGCTTAAATAATCGCGTTTGAATTTTCTATACCGGATTTGATACATT 300
DB 241 ATACTGACGGGAAATGTGATTAATTAATTTGATCTCTTACCACTTAAGCCCATTT 300
OY 301 GAAATCGAGAAATTTTCGAGAAATTCGGCATTCGTCGCAAGTTTAATTCGCAATCCATT 360
DB 301 GAAATCGTTGAAGTTTCAGGATTAAGTGTGTCTATTCCTGATTAATTCGCAATCCAAAT 360
```

```
OY 361 CAGAAACAGGTTTAAACATTACGTTTCGAGAGACGCTGCTATTGAGCGAGCAGCTTT 420
DB 361 CAAAACAGGTTAGTACATTACCTTCTACCAACATGTTATTAATGATTAATACATTAT 420
OY 421 GCTATCATGTTTGTCTTACTACTTAATT 447
DB 421 ATCATCTTAATTTTATTAATGATTTTATT 447
```

```
RESULT 3
US-10-793-626-3743
; Sequence 3743, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMBERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 3743
; LENGTH: 3240
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-10-793-626-3743
```

```
Query Match 40.7%; Score 182; DB 8; Length 3240;
Best Local Similarity 65.5%; Pred. No. 1.1e-40;
Matches 266; Conservative 0; Mismatches 140; Indels 0; Gaps 0;
```

```
OY 1 ATGTTCCGATCAGATTTTATCATCGACCTAATTTTAAAGTGTACTCTCACTTAATTTT 60
DB 2834 ATGATAGGTTCCAGAAATTAATTTCTCTTATTCGTAAGGTGCTGACTCAATTTGATTTT 2893
OY 61 GCGGAAAAACAGGAGATCGTCCGAGAGACTTGTGTACCGGATTTTAACTGTG 120
DB 2894 GCTGAGAAATTTGGATTAATCCACAGGTTAAGTGTTCACAGTTATTTAGCTTTGATT 2953
OY 121 TTTAATCAGCGGCTTTATTTTACTTGTGTGTCTAGTAGAGCTTCTCACTTAATTC 180
DB 2954 TTTGATCAACCGATCATGTTGTATTCAGATTAATCAATTAAGTCTTAATTTTATC 3013
OY 181 GTGAAATACGGTTTATCCAAATTTATGATTTTGTACGAGCAGAGAAATTCGCTCCANG 240
DB 3014 GTAGCAACGGTATTAAGTAAGTGGTTATTTATATGTAAGAAATTCGCTCCANG 3073
OY 241 CTGATTAACAGGAGATCGCTTAAATAATCGCGTTTGAATTTTCTATACCGGATTTGATACATT 300
DB 3074 ATACTGACGGGAAATGTGATTAATTAATTTGATCTCTTACCACTTAAGCCCATTT 3133
OY 301 GAAATCGAGAAATTTTCGAGAAATTCGGCATTCGTCGCAAGTTTAATTCGCAATCCATT 360
DB 3134 GAAATGTTGAAGTTTCAGGATTAAGTGTGTCTATTCCTGATTAATTCGCAATCCAAAT 3193
OY 361 CAGAAACAGGTTTAAACATTACGTTTCGAGAGACGCTGCTATTGA 406
DB 3194 CAAAACAGGTTAGTACATTACCTTCTACCAACATGTTATTAA 3239
```

```
RESULT 4
US-10-793-626-4197/c
; Sequence 4197, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMBERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS480US
```

CURRENT APPLICATION NUMBER: US/10/793,626
 CURRENT FILING DATE: 2004-03-04
 PRIOR APPLICATION NUMBER: 60/164,258
 PRIOR FILING DATE: 1999-11-09
 NUMBER OF SEQ ID NOS: 4472
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 4197
 LENGTH: 3029
 TYPE: DNA
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Description of Artificial Sequence: synthetic
 OTHER INFORMATION: nucleic acid sequence
 US-10-793-626-4197

Query Match 10.2%; Score 45.6; DB 8; Length 3029;
 Best Local Similarity 62.1%; Pred. No. 0.011;
 Matches 72; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

Qy 332 TCGTCCAGTTTAAATGCAATACCAATGAGAAACAGTTTAAACATTACGTTGCGAA 391
 Db 3023 TCATTCCTGGTATATATGGAATACAAATCAAAACAGGTGATGCACTTCTTA 2964
 Qy 392 GCAGCTGTATGAGCGGAGGACCTTGTATCATGTTTGTACTACTAATT 447
 Db 2963 CAAACATGTATTAACATGATATATATATATATATATATATATATATATAT 2908

RESULT 5
 US-11-181-234-1
 Sequence 1, Application US/1181234
 Publication No. US20060021075A1
 GENERAL INFORMATION:
 APPLICANT: WANG, CHYUNG-RU
 TITLE OF INVENTION: GROUP 1 CD1 TRANSGENIC MICE AND THEIR
 TITLE OF INVENTION: USES
 FILE REFERENCE: 21117.0001U2
 CURRENT APPLICATION NUMBER: US/11/181,234
 CURRENT FILING DATE: 2005-07-14
 PRIOR APPLICATION NUMBER: 60/588,192
 PRIOR FILING DATE: 2004-07-15
 NUMBER OF SEQ ID NOS: 7
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 1
 LENGTH: 168753
 TYPE: DNA
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Description of Artificial Sequence: note =
 OTHER INFORMATION: Synthetic Construct
 US-11-181-234-1

Query Match 8.5%; Score 37.8; DB 12; Length 168753;
 Best Local Similarity 49.3%; Pred. No. 6.9; 102; Indels 0; Gaps 0;
 Matches 99; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

Qy 15 TTTATACATGCACTATTTAGTGTACTACGTTTAAATTTGGCGAAAAACAG 74
 Db 91496 TTATATACATTTGCTGCTGCTAGGTTGCTTAAAGTTTCAAGTAAACAAACA 91555
 Qy 75 GATCGTCCGCGAGACCTGTTGTACCGGATATTTAGACCTGTGTTTAAACGCGGT 134
 Db 91556 CCGCGACCTTTGAATGCTGATATATGTTGATTCCTTCGCGAGGAGCTGGC 91615
 Qy 135 CTTTATTTACTGTTTGTAGTACGCTTGCATCTATGATATGCGTAAACGTTT 194
 Db 91616 ACTTCTTTTCTTTGGAATTTGTGCAATTTGATGAGGTTTATATATATATAT 91675
 Qy 195 ATCCAAATTATGATTTGTA 215
 Db 91676 TTTTATATATATATATATATA 91696

RESULT 6
 US-09-925-065A-307190
 Sequence 307190, Application US/09925065A
 Publication No. US20040181048A1
 GENERAL INFORMATION:
 APPLICANT: Wang, David G.
 TITLE OF INVENTION: Identification and Mapping of Single
 TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
 FILE REFERENCE: 108827.135
 CURRENT APPLICATION NUMBER: US/09/925,065A
 CURRENT FILING DATE: 2001-08-08
 PRIOR APPLICATION NUMBER: US 60/243,096
 PRIOR FILING DATE: 2000-10-24
 PRIOR APPLICATION NUMBER: US 60/252,147
 PRIOR FILING DATE: 2000-11-20
 PRIOR APPLICATION NUMBER: US 60/250,092
 PRIOR FILING DATE: 2000-11-30
 PRIOR APPLICATION NUMBER: US 60/261,766
 PRIOR FILING DATE: 2001-01-16
 PRIOR APPLICATION NUMBER: US 60/289,846
 PRIOR FILING DATE: 2001-05-09
 NUMBER OF SEQ ID NOS: 957086
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 307190
 LENGTH: 389
 TYPE: DNA
 ORGANISM: Homo sapiens
 US-09-925-065A-307190

Query Match 7.5%; Score 33.6; DB 6; Length 389;
 Best Local Similarity 52.1%; Pred. No. 12;
 Matches 75; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

Qy 65 AAAAAACAGGATCGTCCGCGAGACTTGTGTTACCGGATATTTAGACTGTGTTTA 124
 Db 70 AAAATCTGTATTTAGTATGAGCCACTGTGATCAATATCTTAGGCTCTTGAGATA 129
 Qy 125 ATGACCCGCTCTTATTTTACTGTTTGTAGAGCTGCTCACTTATGTTATGTA 184
 Db 130 ATTCTCTGAGCTTCTGATTAACACTGCTGCTTCACTTGCACCTTATGTTATGAGA 189
 Qy 185 AATACGTTTATCAAAATTATGA 208
 Db 190 CAGCTTCTTATCAATTTTATGA 213

RESULT 7
 US-09-925-065A-297021/c
 Sequence 297021, Application US/09925065A
 Publication No. US20040181048A1
 GENERAL INFORMATION:
 APPLICANT: Wang, David G.
 TITLE OF INVENTION: Identification and Mapping of Single
 TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
 FILE REFERENCE: 108827.135
 CURRENT APPLICATION NUMBER: US/09/925,065A
 CURRENT FILING DATE: 2001-08-08
 PRIOR APPLICATION NUMBER: US 60/243,096
 PRIOR FILING DATE: 2000-10-24
 PRIOR APPLICATION NUMBER: US 60/252,147
 PRIOR FILING DATE: 2000-11-20
 PRIOR APPLICATION NUMBER: US 60/250,092
 PRIOR FILING DATE: 2000-11-30
 PRIOR APPLICATION NUMBER: US 60/261,766
 PRIOR FILING DATE: 2001-01-16
 PRIOR APPLICATION NUMBER: US 60/289,846
 PRIOR FILING DATE: 2001-05-09
 NUMBER OF SEQ ID NOS: 957086
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 297021
 LENGTH: 498
 TYPE: DNA
 ORGANISM: Homo sapiens

US-09-925-065A-297021

Query Match 7.5%; Score 33.4; DB 6; Length 498;
Best Local Similarity 52.5%; Pred. No. 14;
Matches 73; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 92 TTGTTTACCGGATTTAGACCTTGTATTAACCGGCTTTATTTTACTTGT 151
DB 308 TTGTCCTACCAATCAGCTGACCTTATTTCTCTATCTAAGTATTTTCTTCTT 249
QY 152 TGTATGAGCTTGTCTCATATATGATGAAATACGTTTATCCAAATTTATGAT 211
DB 248 TCTTTTATTTAGTTTCTTATAGCTTATGAGAGTATTCCTTATCCAAATTTAT 189
QY 212 TGTACGACGAGAAAT 230
DB 188 TGTATGATTTAGAAATAT 170

RESULT 8

US-09-925-065A-780420/c
; Sequence 780420, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:

; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925.065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 780420
; LENGTH: 528
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-780420

Query Match 7.5%; Score 33.4; DB 6; Length 528;
Best Local Similarity 48.7%; Pred. No. 15;
Matches 91; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 61 GCGGAAAAACAGGATCGTCCGACGACTTGTGACCGGATTTAGACTGTG 120
DB 500 GAGGAGAGAGAGAGAGCTCACAGCAGGACGCTGCTGCGATCTCCCTTTGA 441
QY 121 TTTATGAGCCGCTTATTTTACTTGTGTTGCTAGAGAGCTTCTCACTATGATC 180
DB 440 TTTAAGCAGACATTTATTTTGTGCTCGATTAATGATTTTGGATTTATGTAATC 381
QY 181 GTGAATACGCTTATTCAAATTTATGATTTTGTACGAGCAGAAATTCGCTCCATG 240
DB 380 CAAAACATTTTTCGCAAAATGACAAAGTGGTTAGAAAATCTGAAATTCCTACAGT 321
QY 241 CTGATTA 247
DB 320 AAGATGA 314

RESULT 9

US-09-925-065A-804269/c
; Sequence 804269, Application US/09925065A
; Publication No. US20040181048A1

GENERAL INFORMATION:

; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925.065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 804269
; LENGTH: 530
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-804269

Query Match 7.5%; Score 33.4; DB 6; Length 530;
Best Local Similarity 48.7%; Pred. No. 15;
Matches 91; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 61 GCGGAAAAACAGGATCGTCCGACGACTTGTGACCGGATTTAGACTGTG 120
DB 500 GAGGAGAGAGAGAGAGCTCACAGCAGGACGCTGCTGCGATCTCCCTTTGA 441
QY 121 TTTATGAGCCGCTTATTTTACTTGTGTTGCTAGAGAGCTTCTCACTATGATC 180
DB 440 TTTAAGCAGACATTTATTTTGTGCTCGATTAATGATTTTGGATTTATGTAATC 381
QY 181 GTGAATACGCTTATTCAAATTTATGATTTTGTACGAGCAGAAATTCGCTCCATG 240
DB 380 CAAAACATTTTTCGCAAAATGACAAAGTGGTTAGAAAATCTGAAATTCCTACAGT 321
QY 241 CTGATTA 247
DB 320 AAGATGA 314

RESULT 10

US-09-925-065A-547674
; Sequence 547674, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:

; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925.065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 547674
; LENGTH: 569
; TYPE: DNA
; ORGANISM: Homo sapiens

US-09-925-065A-547674

Query Match 7.5%; Score 33.4; DB 6; Length 569;
Best Local Similarity 48.7%; Pred. No. 15;
Matches 91; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

116 TTGCTTTAAACAGCCGCTTTATTTACTGTTTGTGCTGAGCTGCTCACTTANG 175
|||
305 TTGACTGACAGCTTTTGATGTGTTCTCCAAATTTCTAGTTGCTCCACACAAAT 364
|||
176 TTATCGTGAATAGCGTTTATCCAAATTTATGATTTTGACGACGACGAAATTCGCTG 235
|||
365 TTGACAAAGCTTTCTTTTACGACAGTTATTTTAAATCACTTTTCAAAATTTCAACT 424
|||
236 CCATGCTGATTAACAGGATCGTCTTAAATGCGTTGATTTTCTATPACCGATTTGAC 295
|||
425 TAGTTTCAAAACATTCACCTTCAAACTTTGTGAAGAGATTAACACTTTTGTG 484
|||
296 CATTGGA 302
|||
485 CATTTAA 491
|||

RESULT 11

US-09-925-065A-547675
Sequence 547675, Application US/09925065A
Publication No. US20040181048A1

GENERAL INFORMATION:

APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single
FILE REFERENCE: 108827.135
CURRENT APPLICATION NUMBER: US/09/925,065A
CURRENT FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: US 60/243,096
PRIOR FILING DATE: 2000-10-24
PRIOR APPLICATION NUMBER: US 60/252,147
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/250,092
PRIOR FILING DATE: 2000-11-30
PRIOR APPLICATION NUMBER: US 60/261,766
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/289,846
PRIOR FILING DATE: 2001-05-09
NUMBER OF SEQ ID NOS: 957086
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 547675
LENGTH: 569
TYPE: DNA
ORGANISM: Homo sapiens
US-09-925-065A-547675

Query Match 7.5%; Score 33.4; DB 6; Length 569;
Best Local Similarity 48.7%; Pred. No. 15;
Matches 91; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

116 TTGCTTTAAACAGCCGCTTTATTTACTGTTTGTGCTGAGCTGCTCACTTANG 175
|||
305 TTGACTGACAGCTTTTGATGTGTTCTCCAAATTTCTAGTTGCTCCACACAAAT 364
|||
176 TTATCGTGAATAGCGTTTATCCAAATTTATGATTTTGACGACGACGAAATTCGCTG 235
|||
365 TTGACAAAGCTTTCTTTTACGACAGTTATTTTAAATCACTTTTCAAAATTTCAACT 424
|||
236 CCATGCTGATTAACAGGATCGTCTTAAATGCGTTGATTTTCTATPACCGATTTGAC 295
|||
425 TAGTTTCAAAACATTCACCTTCAAACTTTGTGAAGAGATTAACACTTTTGTG 484
|||
296 CATTGGA 302
|||
485 CATTTAA 491
|||

RESULT 12

US-09-925-065A-418523
Sequence 418523, Application US/09925065A
Publication No. US20040181048A1

GENERAL INFORMATION:

APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single
FILE REFERENCE: 108827.135
CURRENT APPLICATION NUMBER: US/09/925,065A
CURRENT FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: US 60/243,096
PRIOR FILING DATE: 2000-10-24
PRIOR APPLICATION NUMBER: US 60/252,147
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/250,092
PRIOR FILING DATE: 2000-11-30
PRIOR APPLICATION NUMBER: US 60/261,766
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/289,846
PRIOR FILING DATE: 2001-05-09
NUMBER OF SEQ ID NOS: 957086
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 418523
LENGTH: 611
TYPE: DNA
ORGANISM: Homo sapiens
US-09-925-065A-418523

Query Match 7.5%; Score 33.4; DB 6; Length 611;
Best Local Similarity 52.5%; Pred. No. 15;
Matches 73; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

92 TTGTTGACCGGATATTAGACTTGTGTTATACACCGGCTTTATTTACTGTT 151
|||
194 TTGCTCAACATCACTAGACTTATTTCTTAAAGTATTTTCTTTCTTCT 253
|||
152 TGCTAGTACGCTGCTACTATGTTATCGGAATAGGTTTATCCAAATTTATGATT 211
|||
254 TCTTTTAAATTAATGTTTCTTAGGTTTCAAGAGTACTTTATCCAAATTTATTAAT 313
|||
212 TGACGACGACGAAATTT 230
|||
314 TGTATGATTTTGAATTAAT 332
|||

RESULT 13

US-09-925-065A-418524
Sequence 418524, Application US/09925065A
Publication No. US20040181048A1

GENERAL INFORMATION:

APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single
FILE REFERENCE: 108827.135
CURRENT APPLICATION NUMBER: US/09/925,065A
CURRENT FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: US 60/243,096
PRIOR FILING DATE: 2000-10-24
PRIOR APPLICATION NUMBER: US 60/252,147
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/250,092
PRIOR FILING DATE: 2000-11-30
PRIOR APPLICATION NUMBER: US 60/261,766
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/289,846
PRIOR FILING DATE: 2001-05-09
NUMBER OF SEQ ID NOS: 957086
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 418524
LENGTH: 611
TYPE: DNA
ORGANISM: Homo sapiens

US-09-925-065A-418524

Query Match 7.5%: Score 33.4; DB 6; Length 611;
Best Local Similarity 52.5%: Pred. No. 15;
Matches 73; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

Qy 92 TTGTTGACCGGATATTAGACTTGTGTTAATCAGCCGCTCTTAATTTACTTGT 151
Db 194 TTGTCATCAATCACTAGACCTTAATTTCTCTATCTAAGATATTTCTTCTTCT 253
Qy 152 TGTAGTAGCTTGTCTACTATATGTAATCGAATACGCTTATCCAAATTTATGATT 211
Db 254 TCTTTTAATTTAGTTTCTTAGGTTTCATGAGTAGCTTATCCAAATTTAATAT 313
Qy 212 TGTACGACGAGAAATT 230
Db 314 TGTATGATTTAGAAATAT 332

RESULT 14

US-09-925-065A-692556

; Sequence 692556, Application US/09925065A
; Publication No. US20040181048A1

; GENERAL INFORMATION:

; APPLICANT: Wang, David G.

; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome

; FILE REFERENCE: 108827.135

; CURRENT APPLICATION NUMBER: US/09/925,065A

; PRIOR FILING DATE: 2001-08-08

; PRIOR APPLICATION NUMBER: US 60/243,096

; PRIOR FILING DATE: 2000-10-24

; PRIOR APPLICATION NUMBER: US 60/252,147

; PRIOR FILING DATE: 2000-11-20

; PRIOR APPLICATION NUMBER: US 60/250,092

; PRIOR FILING DATE: 2000-11-30

; PRIOR APPLICATION NUMBER: US 60/261,766

; PRIOR FILING DATE: 2001-01-16

; PRIOR APPLICATION NUMBER: US 60/289,846

; PRIOR FILING DATE: 2001-05-09

; NUMBER OF SEQ ID NOS: 957086

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 692556

; LENGTH: 1524

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-925-065A-692556

Query Match 7.5%: Score 33.4; DB 6; Length 1524;
Best Local Similarity 52.5%: Pred. No. 21;
Matches 73; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

Qy 92 TTGTTGACCGGATATTAGACTTGTGTTAATCAGCCGCTCTTAATTTACTTGT 151
Db 221 TTGTCATCAATCACTAGACCTTAATTTCTCTATCTAAGATATTTCTTCTTCT 280
Qy 152 TGTAGTAGCTTGTCTACTATATGTAATCGAATACGCTTATCCAAATTTATGATT 211
Db 281 TCTTTTAATTTAGTTTCTTAGGTTTCATGAGTAGCTTATCCAAATTTAATAT 340
Qy 212 TGTACGACGAGAAATT 230
Db 341 TGTATGATTTAGAAATAT 359

RESULT 15

US-09-925-065A-692557

; Sequence 692557, Application US/09925065A
; Publication No. US20040181048A1

; GENERAL INFORMATION:

; APPLICANT: Wang, David G.

; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome

; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A

; PRIOR FILING DATE: 2001-08-08

; PRIOR APPLICATION NUMBER: US 60/243,096

; PRIOR FILING DATE: 2000-10-24

; PRIOR APPLICATION NUMBER: US 60/252,147

; PRIOR FILING DATE: 2000-11-20

; PRIOR APPLICATION NUMBER: US 60/250,092

; PRIOR FILING DATE: 2000-11-30

; PRIOR APPLICATION NUMBER: US 60/261,766

; PRIOR FILING DATE: 2001-01-16

; PRIOR APPLICATION NUMBER: US 60/289,846

; PRIOR FILING DATE: 2001-05-09

; NUMBER OF SEQ ID NOS: 957086

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 692557

; LENGTH: 1524

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-925-065A-692557

Query Match 7.5%: Score 33.4; DB 6; Length 1524;
Best Local Similarity 52.5%: Pred. No. 21;
Matches 73; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

Qy 92 TTGTTGACCGGATATTAGACTTGTGTTAATCAGCCGCTCTTAATTTACTTGT 151
Db 221 TTGTCATCAATCACTAGACCTTAATTTCTCTATCTAAGATATTTCTTCTTCT 280
Qy 152 TGTAGTAGCTTGTCTACTATATGTAATCGAATACGCTTATCCAAATTTATGATT 211
Db 281 TCTTTTAATTTAGTTTCTTAGGTTTCATGAGTAGCTTATCCAAATTTAATAT 340
Qy 212 TGTACGACGAGAAATT 230
Db 341 TGTATGATTTAGAAATAT 359

Search completed: February 27, 2006, 11:38:28
Job time : 270.374 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OW nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:34:23 ; Search time 105.186 Seconds
(without alignments)
7553.984 Million cell updates/sec

Title: US-10-789-164-2

Perfect score: 447
Sequence: 1 atgttcgacatgacattata.....tggtgttactactaatt 447

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:*

1: /cgn2_6/prodata/1/ina/1_COMB.seq:*
2: /cgn2_6/prodata/1/ina/5_COMB.seq:*
3: /cgn2_6/prodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/prodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/prodata/1/ina/H_COMB.seq:*
6: /cgn2_6/prodata/1/ina/CCTUS_COMB.seq:*
7: /cgn2_6/prodata/1/ina/PP_COMB.seq:*
8: /cgn2_6/prodata/1/ina/RE_COMB.seq:*
9: /cgn2_6/prodata/1/ina/backfillseq.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	189.4	42.4	453	US-09-710-279-2323	Sequence 2323, Ap
2	189.4	42.4	459	US-09-134-001C-1666	Sequence 1666, Ap
3	182	40.7	3240	US-09-710-279-3743	Sequence 3743, Ap
4	52.8	11.8	198	US-09-134-001C-1489	Sequence 1489, Ap
5	45.6	10.2	3029	US-09-710-279-4197	Sequence 4197, Ap
6	35.2	7.9	4078	US-09-928-847B-43	Sequence 43, Ap
7	34.4	7.7	534	US-09-593-995-6	Sequence 6, Appl
8	34.4	7.7	534	US-10-139-667-6	Sequence 6, Appl
9	34.4	7.7	51259	US-08-781-891-209	Sequence 209, App
10	34.4	7.7	51259	US-09-618-166-209	Sequence 209, App
11	34.2	7.7	1266	US-09-949-016-15338	Sequence 5532, A
12	34.2	7.7	1266	US-08-961-527-137	Sequence 137, App
13	34.2	7.7	260247	US-09-949-016-13358	Sequence 13358, A
14	33.6	7.5	96109	US-09-596-002-35	Sequence 35, App
15	33.2	7.4	501	US-09-248-796A-434	Sequence 434, App
16	33.2	7.4	524032	US-09-949-016-16928	Sequence 16928, A
17	33.2	7.4	524032	US-09-949-016-16928	Sequence 16928, A
18	33.2	7.4	524032	US-09-949-016-16930	Sequence 16930, A
19	33.2	7.4	524032	US-09-949-016-16931	Sequence 16931, A
20	33.2	7.4	529885	US-09-949-016-14340	Sequence 14340, A
21	33.2	7.4	529885	US-09-949-016-14341	Sequence 14341, A
22	33.2	7.4	529885	US-09-949-016-14342	Sequence 14342, A
23	33.2	7.4	529885	US-09-949-016-14343	Sequence 14343, A
24	33.2	7.4	529885	US-09-949-016-14344	Sequence 14344, A

25	33.2	7.4	529885	3	US-09-949-016-14345	Sequence 14345, A
26	33.2	7.4	529885	3	US-09-949-016-14346	Sequence 14346, A
27	33.2	7.4	529885	3	US-09-949-016-14347	Sequence 14347, A
28	33	7.4	2468	3	US-09-533-559-23	Sequence 23, Appl
29	33	7.3	6004	3	US-08-961-527-27	Sequence 27, Appl
30	32.8	7.3	474	3	US-09-621-976-18033	Sequence 18033, A
31	32.8	7.3	85665	3	US-09-949-016-17345	Sequence 17345, A
32	32.6	7.3	1106	3	US-09-107-433-1438	Sequence 1438, Ap
33	32.6	7.3	1347	3	US-09-583-110-1828	Sequence 1828, Ap
34	32.4	7.2	601	3	US-09-949-016-66363	Sequence 86363, A
35	32.4	7.2	1929	3	US-09-248-796A-772	Sequence 772, App
36	32.4	7.2	20441	3	US-09-949-016-14476	Sequence 14476, A
37	32.4	7.2	24645	3	US-09-949-016-14062	Sequence 14062, A
38	32.4	7.2	422592	3	US-09-949-016-14182	Sequence 14182, A
39	32.2	7.2	139552	3	US-09-949-016-15300	Sequence 15300, A
40	32	7.2	601	3	US-09-949-016-47889	Sequence 47889, A
41	32	7.2	601	3	US-09-949-016-47890	Sequence 47890, A
42	32	7.2	601	3	US-09-949-016-94477	Sequence 94477, A
43	32	7.2	1220	2	US-08-940-767-3	Sequence 3, Appl
44	32	7.2	1220	3	US-08-941-155-3	Sequence 3, Appl
45	32	7.2	1220	3	US-09-344-666-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-09-710-279-2323
; Sequence 2323, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS480US
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2323
; LENGTH: 453
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-09-710-279-2323

Query Match	42.4%	Score 189.4	DB 3	Length 453
Best Local Similarity	64.0%	Pred. No. 2.8e-51		
Matches 286	Conservative 0	Mismatches 161	Indels 0	Gaps 0
QY	1	ANGTCGATCAGATTATATACATCGCATATTTAGTGTACTACTAGTTATTTT	60	
DB	1	ATCATGCTTCAGAAATTAATTTTCCTTATTCGTAAGTGCTGACTAGTTGATATTT	60	
QY	61	GGCGAAAAAACAAGGATCGTCCGCGACAGACTTTGTACCGGATTTTAGACTTGC	120	
DB	61	GCTGAGAAATTTGGATTAATCCAGCAGGATTAAGTGTTCAGGTTATTTAGCTTGAAT	120	
QY	121	TTTAAATCAGCCGCTTTATTTTACTGTTTTCGTAAGGCTGCTCAATATGTTATC	180	
DB	121	TTTAAATCAGCCGCTTTATTTTACTGTTTTCGTAAGGCTGCTCAATATGTTATC	180	
QY	121	TTTAAATCAGCCGCTTTATTTTACTGTTTTCGTAAGGCTGCTCAATATGTTATC	180	
DB	121	TTTAAATCAGCCGCTTTATTTTACTGTTTTCGTAAGGCTGCTCAATATGTTATC	180	
QY	181	GTAAGTACGAGTTCATCAATTTATGTTTGTACGACGAGAAATTCGTCGCATG	240	
DB	181	GTAAGTACGAGTTCATCAATTTATGTTTGTACGAGGAGAAATTCGTCGCATG	240	
QY	241	CTGATACAGGAGTGTCTTAAATTCGCTTTGATTTCTATACCCGATGTGACATTT	300	
DB	241	ATACCTACGAGGATGTGATTAATTTATTTGATCTCTGTACCCATGACCCATTT	300	

Qy	301	GAAATCGAGAAATTTCGAGGAATCGGACATCATCGGCCAGGTTTAATTGCCAAATCCATT	360
Db	301	GAAATGGTGAAGTTTCAGGTATAGTGCTGTGCATTTCCGGTAATTAATTCGCATCAAAAT	360
Qy	361	CAGAACAAGGTTTAAACCATTACGTTCCGGAAGCAGCTGCTAATTGAGCGGACGACCTTT	420
Db	361	CAAAAACAAGGTTGATGCTCATTTACACTTTCTTACACAAGTAATTTAACATGATTTACATAAT	420
Qy	421	GCTATCATGTTTGTTTACTACTTAATT	447
Db	421	AATCATCTTAATTTTATATATAGTTTTAAT	447

```

RESULT 2
US-09-134-001C-1666
: Sequence 1666, Application US/09134001C
: Patent No. 6380370
: GENERAL INFORMATION:
: APPLICANT: Lynn Doucette-Stamm et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
: TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: GTC-007
: CURRENT APPLICATION NUMBER: US/09/134,001C
: CURRENT FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/064,964
: PRIOR FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,779
: PRIOR FILING DATE: 1997-08-14
: NUMBER OF SEQ ID NOS: 5674
: SEQ ID NO 1666
: LENGTH: 459
: TYPE: DNA
: ORGANISM: Staphylococcus epidermidis
: US-09-134-001C-1666

```

Query Match	42.4%	Score 189.4	DB 3	Length 459
Best Local Similarity	64.0%	Pred. No. 2.8e-51		
Matches 286	Conservative	0	Mismatches 161	Indels 0
				Gaps 0

Oy	1	ATGTTCCGATCAGATTATTAACATCCGACCTAAATTTTAAAGTGTAACCTACTAGTTAAATTTT	60
Db	7	ATGATAGGTTCCAGAAATTATATTCTTCCTAATCGTAGGTGTGTACTCAGTTTGATATTT	66
Oy	61	GGCGAAAAAACAGGGATCGTCCCGCAGAGACTGTGTGTAACCGGATATTAGACCTGTG	120
Db	67	GCTGGAAATTTGGGATTAATCCACGAGGGTAGTCGTTCACAGTTATTACCTTTGATT	126
Oy	121	TTTATATCAGCCGGTCTTATATTTACTGTGTGTGCTAGTAGCTGTGCTACTATGTTATC	180
Db	127	TTTGATCAACCGATCATGTGTGTATCAGGATTAATCATATTAGTTGCTTAACCTATTATTTATC	186
Oy	181	GTGAATAACGGTTTATCCAAATTTATGATTTTGTGACGACGAGAAAATTCGTCCGCAATG	240
Db	187	GTAAACAACGGATTAAGTAAGTAGGGTTATTTTAAATGTAGAGAAATAATTCCTCGCCATG	246
Oy	241	CTGATTAACAGGATCGTCTTAATAAAATCCGGTTTGATTTTCTATACCCGATTTGACATTT	300
Db	247	ATACGACGGGAATGGTAGTTAAATTTATATTGTGATCTTGTGATCCCATTTGACCCCATTT	306
Oy	301	GAATTCGCAGAATTTTCGAGGAATCGGCATCATCGTCCAGGTTTAATTCGCAATACATT	360
Db	307	GAATTCGTTGAAGTTTCAGGTATAGGTGTGTGCATTTCCCTGATATTAATTCGAAATCAATT	366
Oy	361	CAGAAACAAGGTTTAACCATTAACGTTCCGAGACGCTGCTATTTGACGGAGGACCTTT	420
Db	367	CAAAAACAAGGTGTGATCATTTACACTTTCTACAACAATGTTATTAACATGATTTACATAT	426
Oy	421	GCTATCATGTTGTTTACTACTTAATT	447
Db	427	ATCATCTTATTTTATATAGTTTATT	453

```

US-09-710-279-3743
Sequence 3743, Application US/09710279
Patent No. 6703492
GENERAL INFORMATION:
APPLICANT: KIMMERLY WILLIAM JOHN
TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
FILE REFERENCE: PU348005
CURRENT APPLICATION NUMBER: US/09/710,279
CURRENT FILING DATE: 2000-11-09
PRIOR APPLICATION NUMBER: 60/164,258
PRIOR FILING DATE: 1999-11-09
NUMBER OF SEQ. ID NOS: 4472
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3743
LENGTH: 3240
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-09-710-279-3743

```

Query Match	192	Score	192	DB 3	length	3240			
Best Local Similarity	65.5%	Pred. No.	1	7e-48					
Matches	266	Conservative	0	Mismatches	140	Indels	0	Gaps	0

Qy	1	ATGTTGGATCAGATTATATACATGGCACTAATTTAGTGTACATCACTGATTAATTTT	60
Db	2834	ATGATAGGTTGAGAAATATATTTTCTCTATTCGTAGGTGTGACTACAGTTTATATTT	2893
Qy	61	GCGGAAAAAACAGGAGATCGTGCCTGGCAGAGACTTGTGTACCGGAGATTTTAAAGACTGTG	120
Db	2894	GCTAGAAATTTGGGATTTAATCCAGACAGGGTTACTCGTCCAGGTTATTATAGCTTTGATT	2953
Qy	121	TTTATACGCCGCTCTTATTTTACTTGTTTGCTAGTGAAGCTTGCACCTTATGTTATC	180
Db	2954	TTTATATCAACCGATCACTGTTGTATCAGTATTAATCAATTAGTCTTAACTTATTTATC	3013
Qy	181	GTGAATATCGGTTATATCCAAATTTATATGATTTTATACGAGACGAGAAAATTCGCTGCATG	240
Db	3014	GTAAGCAACGGTATTTAGTAACTGGGTTATTTTATATGTAGAAAGAAAATTCGCTGCATG	3073
Qy	241	CTGATTAACAGGAGTCGTCTTAAAAATCGCGTTGATTTTCTAATCCGATTTGATCAATTT	300
Db	3074	ATTACTGACGGGAATGGTGATTAATTAATTTATTTGATCTCTTGATCCATGATGACCCCATTT	3133
Qy	301	GAAATCGCAGATTTTCAGAGAAATGGCATCATGTCGCAGGTTTAATTTGCCAATATACATT	360
Db	3134	GAAATGTGTGAAGTTTCAGGATATAGGTGTGTCAATTCCTGGTATATATTTGCGAATATACATT	3193
Qy	361	CAGAAACAGGTTTAAACCATTCGTTGCGAAGACAGCGCTCATTTGA	406
Db	3194	CAAAAACAGGTTGTAGTCATTTACCTTTCTTCAACAATGTTATTTATTA	3239

```

RESULT 4
US-09-134-001C-1489/c
; Sequence 1489, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lymn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-0007
; CURRENT APPLICATION NUMBER: US/09/134, 001C
; PRIOR APPLICATION NUMBER: 1998-08-13
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055, 779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1489
; LENGTH: 198

```


TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1489

Query Match
Best Local Similarity 11.8%; Score 52.8; DB 3; Length 198;
Matches 72; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 1 ATGTCGATCAGATTATACATCGCACTAATTAGGTACTGATCACTTAAATTTT 60
DB 104 ATGATAGTTGAGATTATATTTCTTATTCGAGGTCGCTCAGTTGATATTT 45
QY 61 GCGGAAAAACAGGATCGTCCGCGAGACTGTTTGAACCGG 104
DB 44 GCTGAGAAATTTGGATTATTCACAGAGGTTAGTGTCCAGG 1

RESULT 5

US-09-710-279-4197/c
Sequence 4197, Application US/09710279

PATENT No. 6703492
GENERAL INFORMATION:
APPLICANT: KIMMERLY, WILLIAM JOHN
TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
FILE REFERENCE: P0348005
CURRENT APPLICATION NUMBER: US/09/710,279
CURRENT FILING DATE: 2000-11-09
PRIOR APPLICATION NUMBER: 60/164,258
NUMBER OF SEQ ID NOS: 1999-11-09
SOFTWARE: Patent Ver. 2.1
SEQ ID NO 4197
LENGTH: 3029
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-09-710-279-4197

Query Match
Best Local Similarity 10.2%; Score 45.6; DB 3; Length 3029;
Matches 72; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 332 TCGTCCAGGTTTAAATGCAATACCATTCAGAAACAGGTTTAAACATTACGTTGGAA 391
DB 3023 TCAATCTCGTATATATGGAATACATTAACAAAGGTGATGATTAACACTTTCTA 2964
QY 392 GCACCGCTGCTATTAGCGGAGGACCTTTGCTATCATGTTTGTACTTAAATT 447
DB 2963 CAACAAATGTTATTAACATGATATACATATATCATCTTATTTTATATAGTTTATT 2908

RESULT 6

US-09-928-847B-43
Sequence 43, Application US/09928847B

PATENT No. 6808896
GENERAL INFORMATION:
APPLICANT: No. 6808896ozymes A/S
APPLICANT: Jorgensen, Steen T
APPLICANT: Rasmussen, Michael D
APPLICANT: Andersen, Jens Tonne
TITLE OF INVENTION: Multiple Insertion of Genes
FILE REFERENCE: 10022.204-US
CURRENT APPLICATION NUMBER: US/09/928,847B
CURRENT FILING DATE: 2002-05-06
NUMBER OF SEQ ID NOS: 50
SOFTWARE: Patent version 3.1
SEQ ID NO 43
LENGTH: 4078
TYPE: DNA
ORGANISM: Bacillus licheniformis

US-09-928-847B-43

Query Match
Best Local Similarity 7.9%; Score 35.2; DB 3; Length 4078;
Matches 76; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 203 TTATGATTTGTACGACGAGCAAAATTCGTGTCATGCTGATTAACAGGATGCTCTAA 262
DB 2799 TTACATTTTAACTGATCTTTTCCCGGCTATGACCTGCCGAGAAAGATCCCTTCC 2858
QY 263 AAATCGCTTGAATTTCTATACCCGATTTGACATTTGAAATGCGAATTTGAGAGA 322
DB 2859 AAATCGCTGCGCCCTCCGCAAGAGGTTGAAGGCTTGAAGGAGGATTCATCAATCA 2918
QY 323 TCGCATCATGTCGCCAGTTTA 346
DB 2919 TTCAGTGTGATGAACCTGCTTGA 2942

RESULT 7

US-09-593-995-6/c
Sequence 6, Application US/09593995

PATENT No. 6406888
GENERAL INFORMATION:
APPLICANT: Conklin, Darrell C.
TITLE OF INVENTION: HELICAL CYTOKINE ZALPHA33
FILE REFERENCE: 99-38
CURRENT APPLICATION NUMBER: US/09/593,995
CURRENT FILING DATE: 2000-06-14
PRIOR APPLICATION NUMBER: 60/139,121
NUMBER OF SEQ ID NOS: 10
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 6
LENGTH: 534
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: degenerate sequence
NAME/KEY: misc.feature
LOCATION: (1)...(534)
OTHER INFORMATION: n = A,T,C or G
US-09-593-995-6

Query Match
Best Local Similarity 7.7%; Score 34.4; DB 3; Length 534;
Matches 73; Conservative 44; Mismatches 141; Indels 0; Gaps 0;

QY 106 TATTTAGACTGTGTTTAAATCAGCCGCTTATTTACTGTTTGCTAGAGCTTG 165
DB 510 DATRTCCCADATNARRTNCKNARATGRTCAATDATTGTTNARYTGTGTTTGTG 451
QY 166 CTCACCTATGTTATGCTGAATAATACGTTTATCAATTTATGATTTTGAACGACAGA 225
DB 450 YTTNARYTTTTRTNACTTCGCDATTCCKNCKYTCYTONSNMCGAANKGANGG 391
QY 226 AAATTCGTCGATCGTATTAACAGGATGCTCTTAAATGCGGTTGATTTCTATAC 285
DB 390 NCONGNCNCRTRCTRTTYTNSMNCRCRYCYTCNACRTANGGDATNARYTGYTCNAC 331
QY 286 CCGATTTGACATTTGAATTCGACAGATTTGAGAGAAATGGGATCATGCTGCAGGTTA 345
DB 330 NEGDAATNGRTCCATNCCNCRCAATTTTCRTTCAATTTTCTANACNARKNARYTT 271
QY 346 ATTGCAATACCATTCAG 363
DB 270 NCRRAANRNACNSMAR 253

RESULT 8

US-10-139-667-6/c
Sequence 6, Application US/10139667

Patent No. 6761882
GENERAL INFORMATION:
APPLICANT: Conklin, Darrell C.
APPLICANT: Gao, Zeren
TITLE OF INVENTION: HELICAL CYTOKINE ZALPHA33
FILE REFERENCE: 99-38
CURRENT APPLICATION NUMBER: US/10/139,667
CURRENT FILING DATE: 2002-05-02
PRIOR APPLICATION NUMBER: US/09/593,995
PRIOR FILING DATE: 2000-06-14
PRIOR APPLICATION NUMBER: 60/139,121
PRIOR FILING DATE: 1999-06-14
NUMBER OF SEQ ID NOS: 10
SOFTWARE: Fastseq for Windows Version 3.0
SEQ ID NO 6
LENGTH: 534
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: degenerate sequence
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(534)
OTHER INFORMATION: n = A,T,C or G
US-10-139-667-6

Query Match 7.7%; Score 34.4; DB 3; Length 534;
Best Local Similarity 28.3%; Pred. No. 0.52;
Matches 73; Conservative 44; Mismatches 141; Indels 0; Gaps 0;

QY 106 TATTAGACTTGTGTTATATCAGCCGCTTTATTACTGTTGCTAGTGAAGCTTG 165
DB 510 DATRCCCADATNARTNCKNARTRTCCATDITGTYTNATYGTGTTTGTG 451
QY 166 CTCACCTATGTTATGCGTAATAGCGTTTATCCAAATTATGATTTTGTAGGACGAGA 225
DB 450 YTNARYYTTRTNNACYTCGCDATYTCNCKNCKYCYTCNSWNGCRANCKNGGNG 391
QY 226 AAATTGCTGCATCTGATACAGGATCGCCCTTAAATGCGCTTGTATTTCTATAC 285
DB 390 NCCNGCNCRTCTRTTYTNSMNCRCYCYCNACRTANAGDATNARTYGTGNAC 331
QY 286 CCGATTGTACATTTGAATCGAGATTTGAGGATCGCATCGTGCAGGTTTA 345
DB 330 NGDATTNGRCCATNCCNCCGCAATTTTCRTTRCAATTTTCRTANACNCKNARATT 271
QY 346 ATGGCAATACCATTCAG 363
DB 270 NCKRANARNACNSWNR 253

RESULT 9
US-08-781-891-209
Sequence 209, Application US/08781891
Patent No. 6090620
GENERAL INFORMATION:
APPLICANT: Yu, Ying-Hui
APPLICANT: Yu, Chang-Hu
APPLICANT: Oshima, Junko
APPLICANT: Mulligan, John T.
APPLICANT: Schellenberg, Gerald D.
TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
NUMBER OF SEQUENCES: 209
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781,891
FILING DATE: 27-DEC-1996
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: No. 6090620tenburg Ph.D., Carol
REGISTRATION NUMBER: 39,317
REFERENCE/DOCKET NUMBER: 240052.419
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 209:
SEQUENCE CHARACTERISTICS:
LENGTH: 51259 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-781-891-209

Query Match 7.7%; Score 34.4; DB 3; Length 51259;
Best Local Similarity 51.3%; Pred. No. 4.1;
Matches 80; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 105 ATATTGACCTGTGTTATATCAGCCGCTTTATTACTGTTGCTAGTGAAGCTT 164
DB 25726 ATATTGATTTGTTATGCTCCCTTTTATTTTATTTTATTTATTTGATACGTCTCT 25785
QY 165 GCTCCTATGTTATGCGTAATAGCGTTTATCCAAATTATGATTTTGTAGGACGACG 224
DB 25786 ATGCGCTCTGTTACTCGGTTAGGTTATCATCTGTTGATTTTTCGTAAGAA 25845
QY 225 AAAATTGCTGCCATGCTGATACAGGATCTCT 260
DB 25846 CCAGCTCTGATTTGTTGATTTCTTGTATAGTTCT 25881

RESULT 10
US-09-618-166-209
Sequence 209, Application US/09618166
Patent No. 6583112
GENERAL INFORMATION:
APPLICANT: Yu, Ying-Hui
APPLICANT: Oshima, Junko
APPLICANT: Mulligan, John T.
APPLICANT: Schellenberg, Gerald D.
TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
NUMBER OF SEQUENCES: 209
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed Intellectual Property Law Group
STREET: 701 Fifth Avenue, Suite 6300
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/618,166
FILING DATE: 17-Jul-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Mcmasters, David D.
REGISTRATION NUMBER: 33,963
REFERENCE/DOCKET NUMBER: 240052.419C1
TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 209:
SEQUENCE CHARACTERISTICS:
LENGTH: 51259 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 209:
US-09-618-166-209

Query Match 7.7%; Score 34.4; DB 3; Length 51259;
Best Local Similarity 51.3%; Pred. No. 4.1;
Matches 80; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

105 ATATTAGACCTGTTTAAATCAACCGCTTTATTTACTGTTTCTAGTAGCTT 164
25726 ATATTGATTTTATGCTCCCTTTATTTGATTTTATTTATTTGATAGTCTCT 25785

165 GCTCACTTATGTTATCGTAATACGTTTATCCAAATTTATGATTTGTACGACGAG 224

25786 ATGCCCTCTGTTACTCTGCTAGGTTTATCTATCTGTTGATTTTTCAGAGAA 25845

225 AAAATCGCTGCACGCTGATTAACAGGATGCTCT 260
25846 CCAGCTCCTAGTTTGTGATTTCTTTGATAGTTCT 25881

RESULT 11

US-09-949-016-55321

Sequence 55321, Application US/09949016
Patent No. 6812339

GENERAL INFORMATION:

APPLICANT: VENTER, J. Craig et al.

TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

FILE REFERENCE: CL001307

CURRENT APPLICATION NUMBER: US/09/949,016

CURRENT FILING DATE: 2000-04-14

PRIOR APPLICATION NUMBER: 60/241,755

PRIOR FILING DATE: 2000-10-20

PRIOR APPLICATION NUMBER: 60/237,768

PRIOR FILING DATE: 2000-10-03

PRIOR APPLICATION NUMBER: 60/231,498

PRIOR FILING DATE: 2000-09-08

NUMBER OF SEQ ID NOS: 207012

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO: 55321

LENGTH: 601

TYPE: DNA

ORGANISM: Human

US-09-949-016-55321

Query Match 7.7%; Score 34.2; DB 3; Length 601;
Best Local Similarity 49.2%; Pred. No. 0.64;
Matches 90; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

83 CGGAGAGCTGTTGTAACCGGATATTAGACTGTTGTTTATACGCGGCTTTATTT 142
106 CAGCACCACTTATTAACAGGAGATATATCCCATGCTCATACACCGATTTTATTT 165

143 TACTGTTTGTAGTAGCTGCTCACTTATGTTATCGTAATACGGTTTATCCAAAT 202
166 ACAGACTTAGTTGTTGTTTGTATTCGTGATCTTTTCAAACTCACTATAAGTT 225

203 TTATGATTTTGAACGACGAAATTCGCTGCCATGCTGATTAACAGGATGCTCTTA 262
226 TTAAAGATATTAGCATATGTCATAACATTTCTCATTTGAAAAATCAATGTCATCA 285

263 AAA 265
286 AAA 288

RESULT 12

US-08-961-527-137/c

Sequence 137, Application US/08961527
Patent No. 6420135

GENERAL INFORMATION:

APPLICANT: Charles Kunsch

TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences

NUMBER OF SEQUENCES: 391

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/961,527

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB340P1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8512

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 137:

SEQUENCE CHARACTERISTICS:

LENGTH: 1266 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-08-961-527-137

Query Match 7.7%; Score 34.2; DB 3; Length 1266;
Best Local Similarity 52.4%; Pred. No. 2.5; 68; Indels 0; Gaps 0;
Matches 75; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

121 TTATACAGCGGCTTTATTTTACTGTTTGTCTAGTAGGCTTCTCACTTATGTTATC 180
12235 TTGACGTACTGCTGCTGATATCGGTGTTGTCGACGTAATTTGTTACCTTACGGTC 12176

181 GTGAATACGGTTTATTCAAATTTATGATTTTGAACGACGACGAAATTCGCTGCATG 240
12175 AATACAAAGCTTATACCAATTTGATGCTGCTGCTGTAACCTCTTGATTTTG 12116

241 CTGATTAACAGGATGCTCTTAA 263
12115 GTGATAGCTTGAATGCTCCAGAA 12093

12115 GTGATAGCTTGAATGCTCCAGAA 12093

12115 GTGATAGCTTGAATGCTCCAGAA 12093

12115 GTGATAGCTTGAATGCTCCAGAA 12093

12115 GTGATAGCTTGAATGCTCCAGAA 12093

12115 GTGATAGCTTGAATGCTCCAGAA 12093

12115 GTGATAGCTTGAATGCTCCAGAA 12093

12115 GTGATAGCTTGAATGCTCCAGAA 12093

12115 GTGATAGCTTGAATGCTCCAGAA 12093

12115 GTGATAGCTTGAATGCTCCAGAA 12093

12115 GTGATAGCTTGAATGCTCCAGAA 12093

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:33:05 ; Search time 6563.47 Seconds
(without alignments)
9873.061 Million cell updates/sec

Title: US-10-789-164-3

Perfect score: 1140
Sequence: 1 atgaaaaagaactgagctt.....gtgacaaactaaatcctcaa 1140

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_da:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vl:*
14: gb_hcg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1136.8	99.7	2989	1	AB016245
2	1136.8	99.7	4315	1	AB039950
3	1133.6	99.4	3738	1	AB046355
4	1132	99.3	1143	6	ES0427
5	1132	99.3	3045	6	ES0424
6	1127.2	98.9	8189	1	BS292954
7	1127.2	98.9	200690	1	BSUB0019
8	1122.4	98.5	2775	1	DQ086153
9	540.2	47.4	110000	1	AB017333
10	540.2	47.4	110000	1	CP000002_36
11	332.4	29.2	686	6	AX433330
12	250.4	22.0	94829	1	AE011191
13	250.4	22.0	94830	1	AE017335
14	250.4	22.0	96231	1	AF188935
15	250	21.9	3244	1	BACCAPABC
16	225.6	19.8	488	6	AX433362
17	103.8	9.1	110000	1	CP000092_3
18	82.6	7.2	110000	1	AE017180

C 19	80.8	7.1	110000	1	BA000043_20	Continuation (21 o
C 20	79.4	7.0	110000	1	CP000089_12	Continuation (13 o
C 21	77.8	6.8	249050	1	AL596165	AL596165 Listeria
C 22	77.8	6.8	349980	6	AX417038	AX417038 Sequence
C 23	77.8	6.8	349980	6	AX417041	AX417041 Sequence
C 24	77.8	6.8	349980	6	AX417041	AX417041 Sequence
C 25	74	6.5	885	6	AX414535	AX414535 Sequence
C 26	73.2	6.4	347456	1	AX416380	AX416380 Sequence
C 27	73	6.4	110000	1	AP06840_28	Continuation (29 o
C 28	72.6	6.4	280050	1	AL591975	Continuation (29 o
C 29	72.6	6.4	349980	6	AX641665	AX641665 Sequence
C 30	71	6.2	291073	1	AE017323	AE017323 Listeria
C 31	68.8	6.0	110000	1	AP008934_03	Continuation (4 of
C 32	66.4	5.8	6006	1	BCE17788	AE017270 Bacillus
C 33	66.4	5.8	288437	1	AE017270	Continuation (6 of
C 34	65.8	5.8	110000	1	AP006840_05	Continuation (7 of
C 35	65.8	5.8	326050	1	BA000028_06	Continuation (17 of
C 36	65.2	5.7	110000	1	AE017355_17	Continuation (18 o
C 37	63.8	5.6	110000	1	CP000001_17	Continuation (18 o
C 38	62.2	5.3	110000	1	AE017225_17	Continuation (18 o
C 39	60.4	5.3	110000	1	AE017334_17	Continuation (18 o
C 40	60.4	5.3	110000	1	AE017029	AE017029 Bacillus
C 41	60.4	5.3	290525	1	AP006840_09	Continuation (10 o
C 42	56.6	5.0	110000	1	CP000023_01	Continuation (2 of
C 43	56.2	4.9	110000	1	CP000024_01	Continuation (2 of
C 44	56.2	4.9	110000	1	AB125962	AB125962 Bacillus
C 45	56	4.9	930	1	AB125962	

ALIGNMENTS

RESULT 1
LOCUS AB016245 2989 bp DNA linear BCT 14-OCT-1999
DEFINITION Bacillus subtilis pgsA, pgsB and pgsC genes, complete cds.
ACCESSION AB016245
VERSION AB016245.1 GI:6045071
KEYWORDS pgsA, pgsC, pgsB.
SOURCE Bacillus subtilis
ORGANISM Bacillus subtilis
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

REFERENCE
AUTHORS Ashiuchi,M., Soda,K. and Misono,H.
TITLE A poly-gamma-glutamate synthetic system of Bacillus subtilis IFO 3336: gene cloning and biochemical analysis of poly-gamma-glutamate produced by Escherichia coli clone cells
JOURNAL Biochem. Biophys. Res. Commun. 263 (1), 6-12 (1999)
PUBMED 10486244

REFERENCE
AUTHORS Ashiuchi,M.
TITLE Direct Submission
JOURNAL Submitted (14-JUL-1998) Makoto Ashiuchi, Kochi University, Research Institute of Molecular Genetics; Otsu 200, Monobe, Nankoku, Kochi 783-8502, Japan (E-mail:ashiuchi@img.kochi-u.ac.jp, Tel:0888-64-5215, Fax:0888-64-5109)
FEATURES
source
1..2989
Location/Qualifiers
/organism="Bacillus subtilis"
/mol_type="genomic DNA"
/strain="IFO 3336"
/sub_species="Bacillus natto"
/db_xref="taxon:1423"
1..1182
/gene="pgsB"
1..1182
/gene="pgsB"
/codon_start=1
/transl_table=1
/product="PgsB"
/protein_id="BA85263.1"
/db_xref="GI:6045072"
/translation="WMLLIACAVILVIGLEKRRHQKNIDALPVRVINGIRGKSTV

```

TRLTGILLIAGYKTVGKTGTGTDARMTYMDPEEKPIKRPQSGISEQKEMRETYE
RGANAIVSECAVNDYQIIFQEBELQANIGVINVLIEDHWGPTLDEIAEAFAT
IPNGHLYITDSBYETEEFKKAKERNTKVTIADNSKTDYLRKPEYVVPDPA.SLAL
GVAPALGIDETIARKMLNAPDPDGAARILULIPSEBGEHVNQPAANDASTINTIK
RVKRIGYPTDPIIIMNCRADRVRTQDPADVLPYLEASBLIIGETTERIYVAYES
GKIPADKIHDLREKYSTDEIMELLKSMNHRVYGVGNHGAEBLIEKIHBYKXKOLV
S"
gene
1197..1646
/ gene="pgc"
1197..1646
/ gene="pgc"
/ gene="pgc"
/ codon_start=1
/ trans1_table=1
/ product="Pgsc"
/ protein_id="BA85264.1"
/ db_xref="GI:6045073"
/ translation="MPSDLYIALILGVLSLIPAEKTIYVAGLVPYGLVENDP
VFILLVLVLSLITVIVKYGSKEMILKRRKFAMLITGVILKIAEDPLVIVPEI
AEFRIGIIVPLIANTIQOGKITFGSTLLSGATPAIMFYLLI"
1665..2807
/ gene="pgsa"
1665..2807
/ gene="pgsa"
/ codon_start=1
/ trans1_table=1
/ product="Pgsa"
/ protein_id="BA85265.1"
/ db_xref="GI:6045074"
/ translation="MKKELSPHEKLLKLTQOKKKNKHVIAIPIVFLVFMFMWAG
KAETPKVTVSDVLSAFVDIMMGRVVEVTEQKADSIFQYVEPIFRASDVAGN
FENPTVQKNYKQADKEIHLQTNKESVVLKDMNFTVANSANNHMDVQGMQDTLG
EPAKONLDIVAGVSLSDAKKISYOKNYSIATLIGTPTDVSNGPAKKNTPGLPA
DPEIFIPMISAKKHADIVVVOSSHGQGYDNDPRDROQLAASDADADIVGHHPR
VLEPIEVNGTVIFPSLGNFVFDQGMTRUSALVOYHLKNGKREFEVPTIDHEAT
PAPVKOSLKKOTIIRLTKDSNFAKVEDKLTFTDIDSKLSK"

```

ORIGIN

Query Match 99.7%; Score 1136.8; DB 1; Length 2989;

Best Local Similarity 99.8%; Pred. No. 2.9e-276; Indels 0; Gaps 0;

Matches 1138; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

OY 1 ATGAAAAAGAACTGAGCTTTCATGAAAAAGCTCTAAAGCTGACAAAAAGCAAAAAAAG 60
DB 1665 ATGAAAAAGAACTGAGCTTTCATGAAAAAGCTCTAAAGCTGACAAAAAGCAAAAAAAG 1724
OY 61 AAAACCAATAGCAGTATTTATGCGCATTCGATCGTTTTTGTCTTATGTCCTTTC 120
DB 1725 AAAACCAATAGCAGTATTTATGCGCATTCGATCGTTTTTGTCTTATGTCCTTTC 1784
OY 121 ATGTGGGGGGGAAAGCGGAAAGCGCGAAGTCGTAATTCGACGAGCTACTCTCA 180
DB 1785 ATGTGGGGGGGAAAGCGGAAAGCGCGAAGTCGTAATTCGACGAGCTACTCTCA 1844
OY 181 GCCTCATTTGTAGCGGATATTATGATGGGACGCTATGTTGAAAAAGTAACGGAGCAAAA 240
DB 1845 GCCTCATTTGTAGCGGATATTATGATGGGACGCTATGTTGAAAAAGTAACGGAGCAAAA 1904
OY 241 GGGGCGACAGTATTTTCAATATGTTGAACCGATCTTTAGAGCTCGGATTAATGACGA 300
DB 1905 GGGGCGACAGTATTTTCAATATGTTGAACCGATCTTTAGAGCTCGGATTAATGACGA 1964
OY 301 GGAACCTTTGAAAAACCGGTAACCTATCAAAAGATTTATAACAGCAGATTAANAAGTT 360
DB 1965 GGAACCTTTGAAAAACCGGTAACCTATCAAAAGATTTATAACAGCAGATTAANAAGAT 2024
OY 361 CATCTGCAGACGAATAGGAATCAGTGAAGAGCTTGAAGAGATGAAATTTCACGGTTCTC 420
DB 2025 CATCTGCAGACGAATAGGAATCAGTGAAGAGCTTGAAGAGATGAAATTTCACGGTTCTC 2084
OY 421 AACAGCGCCAAACACGCAATGATTAACGCGCTTCAGGGCAGTAAAGATACGGTTGGA 480
DB 2085 AACAGCGCCAAACACGCAATGATTAACGCGCTTCAGGGCAGTAAAGATACGGTTGGA 2144

```

```

OY 481 GAATTTGCGAAGCAAAACCTTGATATGCTTGAGCGGGATACAGCTTAAGTATGCGAAA 540
DB 2145 GAATTTGCGAAGCAAAACCTTGATATGCTTGAGCGGGATACAGCTTAAGTATGCGAAA 2204
OY 541 AAGAAATTTTGCATCCGAAAGATCAACGGGTAACGATTTGCAACGCTTGCTTTACCGAT 600
DB 2205 AAGAAATTTTGCATCCGAAAGATCAACGGGTAACGATTTGCAACGCTTGCTTTACCGAT 2264
OY 601 GTGTCCGGGAAAGGTTTCGGGCTTAAAGAAATACCGCGGGCGTGTCCCGCAGATCCT 660
DB 2265 GTGTCCGGGAAAGGTTTCGGGCTTAAAGAAATACCGCGGGCGTGTCCCGCAGATCCT 2324
OY 661 GAATCTTCATCCCTATGATTTTCAAGCGAAAAAATCATGCTGACATTTGTTTGTCAG 720
DB 2325 GAATCTTCATCCCTATGATTTTCAAGCGAAAAAATCATGCTGACATTTGTTTGTCAG 2384
OY 721 TCACACTGGGGCCCAAGATATGACATGATTCGAAACGACCGCCAGCGCCAGCTTGCAAGA 780
DB 2385 TCACACTGGGGCCCAAGATATGACATGATTCGAAACGACCGCCAGCGCCAGCTTGCAAGA 2444
OY 781 GCCATGTCTGATGCGGGAGGCGTACATCATGTCGGCCATCATCCGACGCTTTAGAACCG 840
DB 2445 GCCATGTCTGATGCGGGAGGCGTACATCATGTCGGCCATCATCCGACGCTTTAGAACCG 2504
OY 841 ATTAAGATATATTAACGAAACCGTCAATTTTCTACAGCCTCGGCAACTTTGTCTTGAACAA 900
DB 2505 ATTAAGATATATTAACGAAACCGTCAATTTTCTACAGCCTCGGCAACTTTGTCTTGAACAA 2564
OY 901 GCGTGAACGAGACAAAGACAGTGCATCTGCTTCAATATCACTGAAAGAAATATGAAACA 960
DB 2565 GCGTGAACGAGACAAAGACAGTGCATCTGCTTCAATATCACTGAAAGAAATATGAAACA 2624
OY 961 GCGCGCTTTGAAGAGACACGATGATATCATGATGATGATGATGATGATGATGATGATGAT 1020
DB 2625 GCGCGCTTTGAAGAGACACGATGATATCATGATGATGATGATGATGATGATGATGATGAT 2684
OY 1021 GACAGCCTTAAACGAAACCATATTTGCGAATCTGACGAAAGACTAATTTGCTTGG 1080
DB 2685 GACAGCCTTAAACGAAACCATATTTGCGAATCTGACGAAAGACTAATTTGCTTGG 2744
OY 1081 AAGTAGAAGACGGAACCTGACGTTGATATGATGATGATGATGATGATGATGATGATGAT 1140
DB 2745 AAGTAGAAGACGGAACCTGACGTTGATATGATGATGATGATGATGATGATGATGATGAT 2804

```

RESULT 2

AB039950

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1..4315

/organism="Bacillus subtilis"

/mol_type="genomic DNA"

/strain="Asahikawa"

/db_xref="taxon:1423"

AB039950 4315 bp DNA linear BCT 12-SEP-2000

Bacillus subtilis capB, capC, capA, ywC genes, complete cds.

AB039950.1 GI:10119860

capA, capC, capB.

Bacillus subtilis

Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

1 (bases 1 to 4315)

Tran, L.P. and Itoh, Y.

Nucleotide sequence of the capBCA operon in Bacillus subtilis

(nato)

Published Only in Database (2000)

2 (bases 1 to 4315)

Tran, L.P. and Itoh, Y.

Direct Submission

Submitted (08-MAR-2000) Yoshifumi Itoh, National Food Research

Institute, Applied Microbiology; Kamonada 2-1-2, Tsukuba, Ibaraki

305-8642, Japan (E-mail: yosifumi@nifl.affrc.go.jp,

Tel: +81-298-38-8075, Fax: +81-298-38-7996)

Location/Qualifiers

gene 941. .2122
/gene="capB"
941. .2122
/gene="capB"
/codon_start=1
/transl_table=11
/product="CapB"
/protein_id="BAB13484.1"
/db_xref="GI:10119861"
/translation="MWLLIACAVLITVIGLIEKKRHOKNIDALPVRVINGIRGKSTV
TRLTITSLIEAGIKTVGKTTGTGDARMTIMYDPIEEKPIGRKQGNIGEGEWREYTE
RGNAVLISECMAVNDPYQIIFQEBELLQNNIGIVAVFEDHNDVNGPLTDEIAEFNTV
IIPNGHLITVDSYTEYEPFQOKAKERNTVYIADNSKITDELYRFEYVPPDNASL
GVNAGLIDESTAFKQMLNAPDPGAMSLPLISPSBEGHFNQGFANDASSTLNTWK
RVKEIYPTDPLDIINCCADRVDRTOFPANDVLPYASLLILIGETPIYVAEE
GKIPAKLHLDLEYSKTDIEIMELLKKRMNRNVYIGGNHGAAPLIEKHIEYKQGLV
S"
gene 2137. .2586
/gene="capC"
2137. .2586
/gene="capC"
/codon_start=1
/transl_table=11
/product="CapC"
/protein_id="BAB13485.1"
/db_xref="GI:10119862"
/translation="MEGSDLYTALILGYLLSIFAEKTGIVAGLVVGGYLGLENOP
VFLILVILVSLTYIVYKGLSKFMILVGRBFAAMLITGVILKIAPLFLPIYPEEI
AERGGIIVPGILANTIQKGLITTFGSLTLLSGATPAIMFYILI"
gene 2605. .3747
/gene="capA"
2605. .3747
/gene="capA"
/codon_start=1
/transl_table=11
/product="CapA"
/protein_id="BAB13486.1"
/db_xref="GI:10119863"
/translation="MKKELSPBKLLKLTKOOKKTKNKHVLAIPVFLVFAFMAG
KATPPKATYSDVLTASFPVGDIMGRVYEVKTEBQKDSIFQYVEPIFRASDVVAGN
FENPVYQKRYKQADKEIHLQTNKESVYKLKQMTPTVYVNSANNHAMQDQGVDTL
EFKQNLMDIAGVYSLSDAKKI STQKVGAVITATLGTDSGKPAKKTPOVPLR
DPEIFLPISEAKRHADIVVOSHQGEYDNDPNRQOLAKAMSDACADITIVCHHP
VLSEIYVNGTVIFYSLGNFVEQGTWTRDSALVOYHLKNGGTERPEVPIIDIEHAT
PAPVKDQSLKOKTIIHELTFKDSNFAKVEDGKLTFRDIDHSDLKSK"
3762. .3929
/gene="ywcC"
3762. .3929
/gene="ywcC"
/codon_start=1
/transl_table=11
/protein_id="BAB13487.1"
/db_xref="GI:10119864"
/translation="MKFVKAIMPVAVAVLVPFMGSAFKENDQLTDOEKQKIDMENKK
QQDEPVNANK"

ORIGIN	Query Match	99.7%	Score 1136.8	DB 1	length 4315
Best Local Similarity	99.8%	Pred. No. 2.9e-278			
Matches 1138	Conservative	0	Mismatches 2	Indels 0	Gaps 0
Qy	1	ATGAAAAAAAAAGACTGACCTTTCATGAAAAGCTGCTAAAGCTGACAAACAGCAAAAAAAG	60		
Ddb	2605	ATGAAAAAAGAACTGACGCTTTCATGAAAAGCTGCTAAAGCTGACAAACAGCAAAAAAAG	266		
Qy	61	AAAACCAATAGCAAGTATTATTATGCAATTCGATTCGTTTTCCTTAAGTTCGCTTTC	120		
Ddb	2665	AAAACCAATAGCAAGTATTATTATGCAATTCGATTCGTTTTCCTTAAGTTCGCTTTC	272		
Qy	121	ATGTGGGCGGGAAGACCGGAAACGCCGAAAGGTCAAAAAGTATTCGACAGCGTACTCTCA	180		
Ddb	2725	ATGTGGGCGGGAAGACCGGAAACGCCGAAAGGTCAAAAAGTATTCGACAGCGTACTCTCA	278		

OY	18	GGCCATTTTGTAGCCGATATTAATGATGGAGCGCTAATGTTGAAAAGATACGGAGCAAAAA	240
Db	2785	GCTCTATTGTGAGCGATATTATGATGGAGCGCTATCTTGAAAAAGTAAACGAGCAAAAA	2844
OY	241	GGGGCAGACAGATATTTTTCATAATGTTGAACCGATCTTTAGACCTCGATTATGTAGCA	300
Db	2845	GGGGCAGACAGATATTTTTCATAATGTTGAACCGATCTTTAGACCTCGATTATGTAGCA	2904
OY	301	GGAAACTTTTGAABAAACCCGGTAACTCATCAAAAGAAATATATAACAGAGATTAAGAGATT	360
Db	2905	GGAAACTTTTGAABAAACCCGGTAACTCATCAAAAGAGATTTATATAACAGAGATTAAGAGATT	2964
OY	361	CATCTGCAGACGAAATAGAAATCAGTGAAGAGTCTTGAAAGATATGAATTTCAACGGTTCTC	420
Db	2965	CATCTGCAGAGAAATAGAAATCAGTGAAGAGTCTTGAAAGATATGAATTTCAACGGTTCTC	3024
OY	421	AACAGCCGCAACAAACGCAATGAGATTACGGCGTTCAAGGCTATGAAGATATAGCTTTGA	480
Db	3025	AACAGCCCAAAACAAACGCAATGAGATTACGGCGTTCAAGGCTATGAAGATATAGCTTTGA	3084
OY	481	GAATTTGCGAAGCAAAACCTTGATATTCGTTGAGCGGGATATACGCTTAAGTGAATGCGAAA	540
Db	3085	GAATTTGCGAAGCAAAACCTTGATATTCGTTGAGCGGGATATACGCTTAAGTGAATGCGAAA	3144
OY	541	AAGAAAAATTTGTCGACGAGAAAGTCAACGGGGTAAACGATTGCAACGCTTGGCTTTACCGAT	600
Db	3145	AAGAAAAATTTGTCACGAAAGTCAACGGGGTAAACGATTGCAACGCTTGGCTTTACCGAT	3204
OY	601	GTGTCCGGGAAAGGTTTCGGCGCTAAAAAGATATCGCCGGCGTGTCTGCCCGCAGATCTT	660
Db	3205	GTGTCCGGGAAAGGTTTCGGCGCTAAAAAGATATCGCCGGCGTGTCTGCCCGCAGATCTT	3264
OY	661	GAATCTTTCATCCCTATGATTTTCGAAGCCGAAAAACATGCTGACATTTGTTGTGTGACAG	720
Db	3265	GAATCTTTCATCCCTATGATTTTCGAAGCCGAAAAACATGCTGACATTTGTTGTGTGACAG	3324
OY	721	TCACACTGGGGCCAAAGAGTATGACAATGATCAAAACGACGCGCAGCTTGCAAGA	780
Db	3325	TCACACTGGGGCCAAAGAGTATGACAATGATCAAAACGACGCGCAGCTTGCAAGA	3384
OY	781	GCCATGTCTGATGCGGGAGCTGACATCATCGTGGCCATCATCCGCAACGTCTTGAACCG	840
Db	3385	GCCATGTCTGATGCGGGAGCTGACATCATCGTGGCCATCATCCGCAACGTCTTGAACCG	3444
OY	841	ATTGAAGTATATAACGAAACCGTCAATTTTCTACAGCCCTGGCACTTTGCTTGAACCA	900
Db	3445	ATTGAAGTATATAACGAAACCGTCAATTTTCTACAGCCCTGGCACTTTGCTTGAACCA	3504
OY	901	GGCTTGCAGCAAAACAGAGACAGTGCACCTGTTACAGTATCACTGAAAGAAAAATGCAACA	960
Db	3505	GGCTTGCAGCAAAACAGAGACAGTGCACCTGTTACAGTATCACTGAAAGAAAAATGCAACA	3564
OY	961	GGCGCGCTTTGAAGTACACCGATGATATCATGAAGCAGACCTGCGACCTGTGAAAAAA	1020
Db	3565	GGCGCGCTTTGAAGTACACCGATGATATCATGAAGCAGACCTGCGACCTGTGAAAAAA	3624
OY	1021	GACAGCCTTAAACAGAAAAACATTAATTCGCGAACATGACGAAAGACTCTAATTTGGCTTG	1080
Db	3625	GACAGCCTTAAACAGAAAAACATTAATTCGCGAACATGACGAAAGACTCTAATTTGGCTTG	3684
OY	1081	AAAGTGAAGAACGGAAGAACTGACGTTTGTATTTGATCTAATAGTGAACAACTAAAACTTAA	1140
Db	3685	AAAGTGAAGAACGGAAGAACTGACGTTTGTATTTGATCTAATAGTGAACAACTAAAACTTAA	3744

RESULT 3	AB046355	3728 bp	DNA linear	ECT 11-JAN-2002
LOCUS	AB046355			
DEFINITION	Bacillus subtilis ywsc, ywla, ywlb, ywlc genes, complete cds.			
ACCESSION	AB046355			
VERSION	AB046355.1	GI:13591556		
KEYWORDS				
SOURCE	Bacillus subtilis			

ORGANISM Bacillus subtilis
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

REFERENCE 1 Urushibata, Y., Tokuyama, S. and Tahara, Y.
AUTHORS Characterization of the Bacillus subtilis ywsc gene, involved in
TITLE gamma-polyglutamic acid production
JOURNAL J. Bacteriol. 184 (2), 337-343 (2002)
PUBMED 11751809

2 (bases 1 to 3738)
AUTHORS Tahara, Y. and Urushibata, Y.
TITLE Direct Submision
JOURNAL Submitted (18-JUL-2000) Yaeutaka Tahara, Shizuoka University,
Faculty of Agriculture, 836 Ohya, Shizuoka city, Shizuoka 422-8529,
Japan (E-mail: acyctahae@agr.shizuoka.ac.jp,
Tel: 81-54-238-4878 (ex. 7808), Fax: 81-54-237-3028)

FEATURES Location/Qualifiers
Source 1. 3738
/organism="Bacillus subtilis"
/mol_type="genomic DNA"
/db_xref="taxon:1423"
497..1678
/gene="ywsc"
497..1678
/gene="ywsc"
/codon_start=1
/transl_table=1
/protein_id="BAB40948.1"
/db_xref="gi:13591557"
/translation="MWLLIACAVILVIGILEKRRHQKIDALPVRVNINGIRGKSTV
TRLTTGLIEAGYKTVGKTGTDAAMIYWDPEEKPIRKQGPVIGQKMEETVE
RGANAIVSECAVNDYOIFQBELQNIQVIVNVLHMDVMGPTLDEATETAT
IPYGNHLYITDSEYTEFPEKQAKERTTVIADNSKIDETVLRKEVWVPDNLAL
GVAOLGIDETAPFKGMALPDPGAMILPLISPEBGFHVPFPAADASTINIMK
RVKEIGYTDPTIIMNCRAPVRDTPQADVPYIASLIIIGEQEPIVAYAE
GKIPADKIDLEKSTDEIMELKKRMKNRVLYGVNIGHAEPLEIKTHYKQKOLV
S"

1693..2142
/gene="ywca"
1693..2142
/gene="ywca"
/codon_start=1
/transl_table=1
/protein_id="BAB40949.1"
/db_xref="gi:13591558"
/translation="MFGSDLYALILGLVLSLIPAEKTIYAGVIVPGVIGLVFNOP
VFILVLVSLIIYIVYKGYSKPMILYGRKKPAMLTGIVLKAEPFLYPIVPEEI
AEFRGIGIVPEGLANITIQKGLTTFESTLLSGRAFAIMFVYLLI"

2161..3303
/gene="ywcb"
2161..3303
/gene="ywcb"
/codon_start=1
/transl_table=1
/protein_id="BAB40950.1"
/db_xref="gi:13591559"
/translation="MKKELSPHEKLLKLTQOKKKTKNGHFLAIPVFLWFAFMWAG
KAEPKKTVDVDDVVASFVDIMMGRVVEKTBQKQSTFOYERPIFRASDVAGN
EENPTVYKQYKQADKEHLOTNESYVVLDMNFTVANSNNHMDVQVGMKDTGS
EPAIKNDIVGAGTSLDAKKKISTQKNGVITATLGTIVSGKFAKKAPGVPLA
DPELIPMISEAKKHADIVVQSHGQGTDRDNDPRQQLARAMSADADADIIIGHHP
VLEPIEVNGTVIPYSLGNFVFDQGTTRDSALVOYHLKKGGRFPEVTPIDHEAT
PAPVKDLSLKQKTIIRLETKDSNFAMKVEDGLTFDIDHSKLSK"

3318..3485
/gene="ywcc"
3318..3485
/gene="ywcc"
/codon_start=1
/transl_table=1
/protein_id="BAB40951.1"
/db_xref="gi:13591560"
/translation="WKFVKAIMPVVAIVFMFMSAFKENDQLTDQEKQKIDMEKNKI
OOOEPPVANAK"

gene
CDS
ORIGIN

Query Match 99.4%; Score 1133.6; DB 1; Length 3738;
Best Local Similarity 99.6%; Pred. No. 1.9e-277;
Matches 1136; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

1 ATGAAAAAGAACTGACCTTTCATGAAAGCTGTAAAGCTGACAAACGCAAAAAAG 60
2161 ATGAAAAAGAACTGACCTTTCATGAAAGCTGTAAAGCTGACAAACGCAAAAAAG 2220

61 AAACCAATAGCAAGATTTATGATGATGATGATGATGATGATGATGATGATGATG 120
2221 AAACCAATAGCAAGATTTATGATGATGATGATGATGATGATGATGATGATGATG 2280

121 ATGAGCGCGGAAAAAGCGAAACGCCAAGGTCAAAAGTATCTGACGATCTCTCA 180
2281 ATGAGCGCGGAAAAAGCGAAACGCCAAGGTCAAAAGTATCTGACGATCTCTCA 2340

181 GCTCATTTTGTAGCGCATTTATGATGATGATGATGATGATGATGATGATGATGATG 240
2341 GCTCATTTTGTAGCGCATTTATGATGATGATGATGATGATGATGATGATGATGATG 2400

241 GGGGAGACAGATTTTCAATATGATGATGATGATGATGATGATGATGATGATGATG 300
2401 GGGGAGACAGATTTTCAATATGATGATGATGATGATGATGATGATGATGATGATG 2460

301 GGAACCTTTGAAAAACCGGTAACTATCAAAAGATTTATTAACAGAGATTAAGAGATT 360
2461 GGAACCTTTGAAAAACCGGTAACTATCAAAAGATTTATTAACAGAGATTAAGAGATT 2520

361 CATTCGACAGCAATTAAGAAATCGTAAAGTCTTGAAGATGATGATTTTCAAGCTTCTC 420
2521 CATTCGACAGCAATTAAGAAATCGTAAAGTCTTGAAGATGATGATTTTCAAGCTTCTC 2580

421 AACGCCCAACCAACCGCAATGATGATGATGATGATGATGATGATGATGATGATGATG 480
2581 AACGCCCAACCAACCGCAATGATGATGATGATGATGATGATGATGATGATGATGATG 2640

481 GAATTTGCGAAGCAAAACCTTGATATGCTTGAAGCGGATTAACGCTTAAGTATGCGAAA 540
2641 GAATTTGCGAAGCAAAACCTTGATATGCTTGAAGCGGATTAACGCTTAAGTATGCGAAA 2700

541 AAGAAATTTGTCACAGAAAGTCAACGCGGTAAACGATTCGACGCTTGGCTTACCGAT 600
2701 AAGAAATTTGTCACAGAAAGTCAACGCGGTAAACGATTCGACGCTTGGCTTACCGAT 2760

601 GTGTCGGGAAAGGTTTCGGGCTTAAAGAAATACGCGGCGGTGCGCCGCGAGATCCT 660
2761 GTGTCGGGAAAGGTTTCGGGCTTAAAGAAATACGCGGCGGTGCGCCGCGAGATCCT 2820

661 GAATCTTCATCCTTATGATTTTCAAGACGCAAAAAATGCTGACATTTGTTGTGCGAG 720
2821 GAATCTTCATCCTTATGATTTTCAAGACGCAAAAAATGCTGACATTTGTTGTGCGAG 2880

721 TCACATCGGGGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 780
2881 TCACATCGGGGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2940

781 GCCATGCTGATGCGGAGCTGATGATGATGATGATGATGATGATGATGATGATGATG 840
2941 GCCATGCTGATGCGGAGCTGATGATGATGATGATGATGATGATGATGATGATGATG 3000

841 ATTGAAGATTAACGAAACCGTATTTTCAACGCTCGGCACTTTGCTTTTGAACAA 900
3001 ATTGAAGATTAACGAAACCGTATTTTCAACGCTCGGCACTTTGCTTTTGAACAA 3060

901 GGGTGAAGAGAACAGAGACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 960
3061 GGGTGAAGAGAACAGAGACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 3120

961 GGGCGCTTTGAAGTGAAGACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1020
3121 GGGCGCTTTGAAGTGAAGACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 3180

QY		1021	GACGCGCTTAAAGCAGAAAACCATTAATTCGGCAACTGACGAAGAAGCTCTAATTGCGTTGG	1080
Db		3181	GACGCGCTTAAAGCAGAAAACCATTAATTCGGCAACTGACGAAGAAGCTCTAATTGCGTTGG	3240
QY		1081	AAAGTAGAAGACGGAAAACTGACCGTTTGATATGTATGCATAGTGA CAATCTAAATCTTAA	1140
Db		3241	AAAGTAGAAGACGGAAAACTGACCGTTTGATATGTATGCATAGTGA CAATCTAAATCTTAA	3300
RESULT 4				
E50427			1143 bp DNA linear PAT 31-JAN-2002	
LOCUS				
DEFINITION			Process for producing poly-gamma-glutamic acid.	
VERSION			E50427	
KEYWORDS			E50427.1 GI:18629415	
SOURCE			JP 2001017182-A/4.	
ORGANISM			Bacillus subtilis	
REFERENCE			Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.	
AUTHORS			1 (bases 1 to 1143)	
TITLE			Ashihuchi,M., Misono,H. and Soda,K.	
JOURNAL			Process for producing poly-gamma-glutamic acid Patent: JP 2001017182-A 4 23-JAN-2001;	
COMMENT			NAGASE ECO LTD	
OS			Bacillus subtilis (hay bacillus) IFO 3336	
PN			JP 2001017182-A/4	
PD			23-JAN-2001	
PF			09-JUL-1999 JP 1999196335	
PR				
PI			MAKOTO ASHIUCHI, HARUO MISONO, KENJI SODA	
PC			C12N15/09,C08G69/36,C12N1/15,C12N1/19,C12N1/21,C12N5/10, PC	
C12N9/00,C12N9/90,				
PC			C12P13/14,C12N15/00,C12N5/00	
CC				
FH			Key	
FT			source	
FT			location/Qualifiers	
FT			1..1143	
FT			/organism='Bacillus subtilis'	
FT			/mol_type='genomic DNA'	
FT			/db_xref='taxon:1423'	
FEATURES				
Source				
ORIGIN				
Query Match			99.3%; Score 1132; DB 6; Length 1143;	
Best Local Similarity			99.6%; Pred.No.4.7e-277;	
Matches 1135; Conservative			0; Mismatches 5; Indels 0; Gaps 0	
QY		1	ATGAAAAAAAAAGACTGAGCTTTCATGATAAAAGCTGAAGCTGACAAAAACAGCAAAAAAAG	60
Db		1	ATGAAAAAAAAAGACTGAGCTTTCATGATAAAAGCTGTAAGCTGACAAAAACGCAAAAAAAG	60
QY		61	AAACAATAAGCAGATATTTATTCGCCATTCGATCGTTTTGCTTATGTGCGTTTC	120
Db		61	AAACAATAAGCAGATATTTATTCGCCATTCGATCGTTTTGCTTATGTGCGTTTC	120
QY		121	ATGTGGGCGGGAAAAAGCGGAAACGCCGAAGTCAAAACGTATTTCTGACGACGTACTCTCA	180
Db		121	ATGTGGGCGGGAAAAAGCGGAAACGCCGAAGTCAAAACGTATTTCTGACGACGTACTCTCA	180
QY		181	GCCCTATTTGTAGCGCATTTATGTATGATGGAGCCCTATGTGAAAAAAGTAACGAGCAAAAA	240
Db		181	GCCCTATTTGTAGCGCATTTATGTATGATGGAGCCCTATGTGAAAAAAGTAACGAGCAAAAA	240
QY		241	GGGGCAGACAGTATTTTCAATATGTGTAACCGATTTTATGAGCTCGGATTTATGTAGCA	300
Db		241	GGGGCAGACAGTATTTTCAATATGTGTAACCGATTTTATGAGCTCGGATTTATGTAGCA	300
QY		301	GGAACCTTTGAAAAACCCGGTAACCTATCAAAAGAAATTATTAACACAGCAGATAAGAGATT	360
Db		301	GGAACCTTTGAAAAACCCGGTAACCTATCAAAAGAAATTATTAACACAGCAGATAAGAGATT	360

[illegible]

PI MAKOTO ASHICHI, HARUO MISONO, KENJI SODA
 PC C12N5/09, C08G69/36, C12N1/15, C12N1/19, C12N1/21, C12N5/10, PC
 C12N9/00, C12N9/90,
 PC C12P13/14, C12N15/00, C12N5/00
 CC
 FH Key Location/Qualifiers
 FT source 1. .3047
 FT /organism="Bacillus subtilis" (hay bacillus)
 FT IFO 3336⁺
 FT Location/Qualifiers

FEATURES

Source

1. .3045
 /organism="Bacillus subtilis"
 /mol_type="Genomic DNA"
 /db_xref="taxon:1423"

ORIGIN

Query Match 99.3%; Score 1132; DB 6; Length 3045;
 Best Local Similarity 99.6%; Pred. No. 4,9e-277;
 Matches 1135; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

1 ATGAAAAAGAACTGAGCTTCATGAAAGCTGTAAGCTGACAAAACAGAAAAAG 60
 1663 ATGAAAAAGAACTGAGCTTCATGAAAGCTGTAAGCTGACAAAACAGAAAAAG 1742
 61 AAAACCAATAAGCAGTATTTATTTGCGATTCGATCGTATTTGCTTATGCTTTC 120
 1743 AAAACCAATAAGCAGTATTTATTTGCGATTCGATTCGATCGTATTTGCTTATGCTTTC 1802
 121 ATGTGGCGGGAAGAGCGGAAAGCGCGAAGGTCAAAAGTATTTGACGACGATCTTCA 180
 1803 ATGTGGCGGGAAGAGCGGAAAGCGCGAAGGTCAAAAGTATTTGACGACGATCTTCA 1862
 181 GCCTCATTTGAGGGGATTTATGATGGGAGCTGATGTTGAAAAAGTAAACGAGCAAAA 240
 1863 GCCTCATTTGAGGGGATTTATGATGGGAGCTGATGTTGAAAAAGTAAACGAGCAAAA 1922
 241 GGGGCGAGCAGTATTTTCAATATGTTGAACCGATCTTTAGAGCTCGGATTTATGACA 300
 1923 GGGGCGAGCAGTATTTTCAATATGTTGAACCGATCTTTAGAGCTCGGATTTATGACA 1982
 301 GGAACCTTTGAAACCCCGTAACTTATCAAAAGATTTAAACAGCAGATTAAGAGATT 360
 1983 GGAACCTTTGAAACCCCGTAACTTATCAAAAGATTTAAACAGCAGATTAAGAGATT 2042
 361 CATTCGACGAGCAATTAAGCAATGAGTGAAGTCTTGAAGGATTAATTTACGGTTCTC 420
 2043 CATTCGACGAGCAATTAAGCAATGAGTGAAGTCTTGAAGGATTAATTTACGGTTCTC 2102
 421 AACAGCGCAACCAACGACGATGATTAACGGCTTCAAGGCGATGAAGATACGCTTGA 480
 2103 AACAGCGCAACCAACGACGATGATTAACGGCTTCAAGGCGATGAAGATACGCTTGA 2162
 481 GAATTTGCGAAGCAAAACCTTGATATCGTTGAGCGGGATACAGCTTAAAGTATCGAAA 540
 2163 GAATTTGCGAAGCAAAACCTTGATATCGTTGAGCGGGATACAGCTTAAAGTATCGAAA 2222
 541 AAGAAAAATTTGCTACCAAGAGTCAACGGGGTAAAGATTTGCAACGCTTGGCTTACCAT 600
 2223 AAGAAAAATTTGCTACCAAGAGTCAACGGGGTAAAGATTTGCAACGCTTGGCTTACCAT 2282
 601 GTGTCCGGGAAAGGTTTCCGGGCTAAAAAAGATTAACGGGGGCTGCTGCCGACAGATCT 660
 2283 GTGTCCGGGAAAGGTTTCCGGGCTAAAAAAGATTAACGGGGGCTGCTGCCGACAGATCT 2342
 661 GAAATCTTCATCCCTATGATTTCAAGAGGAAAAAATGCTGACATTTGTTGTGAG 720
 2343 GAAATCTTCATCCCTATGATTTCAAGAGGAAAAAATGCTGACATTTGTTGTGAG 2402
 721 TCACACTGGGGCGCAAGATTAACAATGATCCAAACGACCGGCAACGCGCAAGTGTGACA 780
 2403 TCACACTGGGGCGCAAGATTAACAATGATCCAAACGACCGGCAACGCGCAAGTGTGACA 2462
 781 GCCATGTCTGATGCGGAGCTGACATCATCTGCGGCATCATCCGACGCTTAAACCG 840

|||||
 Db 2463 GCCATGTCTGATGCGGAGCTGACATCATCTGCGGCATCATCCGACGCTTTGAAACCG 2522
 841 ATTGAAGATTAATACGAAACCGTCAATTTCTACGACCGCTCGGCACTTTGCTTTGACCA 900
 Db 2523 ATTGAAGATTAATACGAAACCGTCAATTTCTACGACCGCTCGGCACTTTGCTTTGACCA 2582
 901 GGCTGACGAGAACAGACAGACAGTGCACGTGTTGATGATGACCTGAAGAAAAATGAAACA 960
 Db 2583 GGCTGACGAGAACAGACAGACAGTGCACGTGTTGATGATGACCTGAAGAAAAATGAAACA 2642
 961 GGCGGCTTTGAAGTGAACCGATGATATCATTAAGCGACACTGCACTGTGAAAAA 1020
 Db 2643 GGCGGCTTTGAAGTGAACCGATGATATCATTAAGCGACACTGCACTGTGAAAAA 2702
 1021 GACAGCCTTAAACGAAAAACCATTAATTCGGAATCGAAGAAAGCTAAATTTGCTTGG 1080
 Db 2703 GACAGCCTTAAACGAAAAACCATTAATTCGGAATCGAAGAAAGCTAAATTTGCTTGG 2762
 1081 AAAGTGAAGACGAAAAACCTGACCTTTGATATGATGATGATGACAACTAAATCTAAA 1140
 Db 2763 AAAGTGAAGACGAAAAACCTGACCTTTGATATGATGATGATGACAACTAAATCTAAA 2822

RESULT 6
 BS292954 8189 bp DNA linear BCT 02-DEC-2004
 LOCUS B_subtilis ywe [A, B, C, D, E, F, G] and gerBC genes.
 DEFINITION
 VERSION 292954.1 GI:1894764
 KEYWORDS
 yweB gene; yweC gene; yweA gene; yweB gene; yweC gene; yweD gene;
 yweE gene; yweF gene; yweG gene.
 SOURCE
 ORGANISM
 Bacillus subtilis
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 1 (bases 1 to 8189)
 PRESSEAN, E., MOSER, I., BOURSIER, L., CRUZ RAMOS, H. C., DE LA
 FUENTE, V., HULLIO, M. F., LEJONG, C., SCHLEICH, S., SEKOWSKA, A.,
 SONG, B. H., VILLANI, G., KUNST, F., DANCHIN, A. and GLASER, P.
 The Bacillus subtilis genome from gerBC (311 degrees) to 11cR (334
 degrees)
 Microbiology (Reading, Engl.) 143 (PT 10), 3313-3328 (1997)
 9353933
 2 (bases 1 to 8189)
 LEJONG, C., GLASER, P., PRESSEAN, E. and DANCHIN, A.
 Bacillus subtilis gerB downstream sequence
 Unpublished
 3 (bases 1 to 8189)
 GLASER, P.
 Direct Submission
 Submitted (13-MAR-1997) Philippe Glaser, Regulation de l'Expression
 Genetique, Institut Pasteur, 28 Rue du Dr Roux, Paris, 75724,
 FRANCE
 JOURNAL
 PUBMED
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 FEATURES
 source
 Location/Qualifiers
 1. .8189
 /organism="Bacillus subtilis"
 /mol_type="Genomic DNA"
 /strain="168"
 /db_xref="taxon:1423"
 1. .1069
 /gene="yweC"
 <1. .1069
 /gene="yweC"
 /note="product highly similar to Bacillus anthracis CapB
 protein"
 /codon_start=2
 /transl_table=11
 /protein_id="CAB07475.1"
 /db_xref="GI:1894765"
 /db_xref="GOA:Q798J2"
 /db_xref="UniProt/TrEMBL:Q798J2"
 /translation="RGSIVRLRTGLIENGVTVGKTTGTQDARMIYWDTPPEKPIK
 RKPQPRIGGEKVEKRETVKRGANAIVSECAVNPDIYIFQEBILQANIGIVNVLE"

```
gene
RBS
CDS
/translation="DNDMGPTLDEIAEAFATATIPYNGHVLITSEYTEPEFKAKERNTKVIADNSKIT
DEYLRKREYMFPPNAGSLATVAQALGIDETAARKMNAAPDDGAMRILILISPEP
GHFVNGFANDASSLTINWKRKKEIGYPTDPTIIMNCRADVRBKTQOFADVADPYIE
ASELILIGETTEPIVKAEBGKIPADKLDHELYKSTDEIMELKKSMMNRVIYGVGNI
HGAAPLIEKIEHYKQVLS"
join(1069..1075,1084..1533)
/gene="ywca"
1069..1075
/gene="ywca"
1084..1533
/gene="ywca"
/note="product highly similar to Bacillus anthracis CapC
protein"
/codon_start=1
/transl_table=1
/protein_id="CAB07468.1"
/db_xref="GI:1894766"
/db_xref="GOA:P96737"
/db_xref="UniProt/TREMBL:P96737"
/translation="WFGSDLYIALILGLVLSLTPAEKTIYVAGLVVSGYLGVNQP
VEILVILVSLTIVIVKGLSEKMLYGRRRFAMLTGVLKIAFPFLYPTVEFI
AEFRGIGIIVPLMIANTIOKQGLITTFSTLILSGATPAIMFYVLI"
1538..2694
/gene="ywcb"
1538..1544
/gene="ywcb"
1552..2694
/gene="ywcb"
/note="product highly similar to Bacillus anthracis CapA
protein"
/codon_start=1
/transl_table=1
/protein_id="CAB07469.1"
/db_xref="GI:1894767"
/db_xref="UniProt/TREMBL:P96738"
/translation="MKKELSFHEKLKILKQOKKKTKNHVFIPIPIVPIAFMFWAG
KAETPRKVTYSDVTSASFVGDIMMGRYVEKTEQKADSIQIYVEPIPRADYVAGN
FENPVYQKNYKQADKEIHLQTNKESVYKLMNFTVLNSANNHNDIGVQMKQTLG
BEAKQNLIDVAGYSLSDAKKISYQKVNQVITATLGTVDGSKGPAKKNTPGVLP
DEIFIPMISEAKKHADIVVQSHQOEYNDPNRQRLARMSDAGKADIIIGHHPH
VLEPIEYVNGTIVFISLGNFVFDQGMTTRDSALVOYHLKKNGTGRFVETPIDHEAT
-PAPVKKDSLKQKTIIRLETKDSNFMKVEDGKLTIDHSDLKSK"
2697..2914
/gene="ywcc"
2697..2702
/gene="ywcc"
2709..2876
/gene="ywcc"
/codon_start=1
/transl_table=1
/product="unknown"
/protein_id="CAB07470.1"
/db_xref="GI:1894768"
/db_xref="UniProt/Swiss-Prot:P96739"
/translation="MKFVKAIWPFVAVALVFMMSAFKNDQLTDEQKIDMENANKI
QQQEBPVNANK"
2880..2914
/gene="ywcc"
3000..4258
/gene="ywcd"
3000..3007
/gene="ywcd"
3017..4258
/gene="ywcd"
/note="product similar to E. coli NUPC protein and to
Listeria species pto-related proteins"
/codon_start=1
/transl_table=1
/protein_id="CAB07476.1"
/db_xref="GI:1894769"
/db_xref="GOA:P96740"
/db_xref="UniProt/Swiss-Prot:P96740"
/translation="MNTLANWKKFLVAVITCLVPIWTKAETAEADTSELSIVSEAK
```

```
terminator
gene
CDS
/translation="NLIGVQYKGGTEPEKGFDPBSGLIQVFSKADIHLPSSVNDQYKIGTAVPENIKPGD
ILPFKKGSGTGYVTHDALYIGDQWVHSTQSKVILITNYKSSYWGSGYIGARILAA
DPADADYVQAEAKYGVIPVPGSGTSPSGFDGSGVQVYFOALGIYLRSAEQM
AVGKAPAPQINKREDVVPFNSNTYKGTISHAGIYAGARPIQASSEKVTLSYEDW
KSKWTGIRFNDNLTPAKENPLVSEATTIYGEVPIKQGVPTPEPDRAGFOVYQYGA
AGISLPVATYQVNAQYKIEKADLPKPDIVFPQSTLNPSTIYTGNGVAVHTLSNGVT
ITNNNTSTYWKDKYAGSIRVO"
4259..4293
complement(4292..5170)
/gene="ywcf"
complement(4292..5152)
/gene="ywcf"
/note="product similar to Bacillus subtilis Yxh and Yce
proteins and to E. coli YidA protein"
/codon_start=1
/transl_table=1
/protein_id="CAB07471.1"
/db_xref="GI:1894770"
/db_xref="GOA:P96741"
/db_xref="UniProt/TREMBL:P96741"
/translation="MKCIAIDLTLLKESVISAENREAIKRAVDAGILVTICGRA
TFDKALDDIDITPIANGSTIHDTCRILSRILMOAGKAIADYLVLSKNITFEVY
TDDHILSPFDEAKLHAEIDTLKSANPEQTDILMOGANTOPKQGITPIRHISVPD
GGENIYKLCIFSPMDKLOKAKELKHKKLAQTSQGHITELIPASSGKRALTKLA
DIVGIETQDIYAIGDSPDLSPFVAGHRIAMENAIDELKKEKSTFTVKSNDENGVAIF
IDOLLSQYA"
complement(5163..5170)
/gene="ywcb"
5277..5305
/gene="ywcf"
5277..5305
/gene="ywcf"
complement(5308..6291)
/pseudo
complement(5308..6276)
/gene="ywcf"
/note="an in frame TGA stop codon is found at position
5599 to 5601
translation similar to Bacillus subtilis LyrC protein"
/pseudo
/codon_start=1
/transl_table=1
complement(6285..6291)
/gene="ywcf"
6594..8112
/gene="ywcf"
6594..6601
/gene="ywcf"
6609..7982
/gene="ywcf"
/note="product highly similar to metabolite transport
proteins"
/codon_start=1
/transl_table=1
/protein_id="CAB07473.1"
/db_xref="GI:1894771"
/db_xref="GOA:P96742"
/db_xref="UniProt/TREMBL:P96742"
/translation="MKKOSNIWLVPFGALGALYGDYGVISGAILFMKKEBGLNAFT
BGLVVSILVGAIIIGSAGAKLDRFGRKKAIMAAALFCIGLVALAPNGWMLF
RIILGLANGSTTIYPIVLSIELAPKHKGALSALNOMITVGLISYIVNYIADAEA
WRMWGLAAPSLSLILIGLFWPESPRLFTNGESSKAKITIEKRGTDIOEIHDI
KSAEKQDEBGLKELFDPWVRPLIAGLGLAQFTGNVTIITYAKFTTNQFNGSA
SILGTGIGTAVNLMLVAIKIIDIKRPLLPFNAGVNSLIYALAVLFPNTPA
ASMTVTICGVIVFAVSMGCVWMMPELPLPVRGIGTGVSTMLAHVGLIYSLT
VILMEATIGISLFLIYALIGAMFLFRFKVTEKRSLEBIEBDLDRKNGQGAAG
KOQIVGT"
7981..8112
/gene="ywcf"
complement(8019..8189)
/gene="gerBC"
complement(8019..>8189)
```

```

/gene="gezBC"
/codon_start=1
/transl_table=1
/product="gezBC"
/protein_id="CAB07474.1"
/db_xref="GI:1894772"
/db_xref="GOA:P96743"

Query Match      98.9%   Score 1127.2; DB 1; Length 8189;
Beet Local Similarity 99.3%   Pred. No. 8.4e-276;
Matches 1132; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Oy 1 ATGAAAAAGAACTGAGCTTTTCATGAAAGCTGCTAAAGCTGACAAAACAGCAAAAAAG 60
Db 1552 ATGAAAAAGAACTGAGCTTTTCATGAAAGCTGCTAAAGCTGACAAAACAGCAAAAAAG 1611

Oy 61 AAAACCAATAGCAGCTATTTATGCGCATTCGGATCGTTTTGTCCTTATGTTGCTTTC 120
Db 1612 AAAACCAATAGCAGCTATTTATGCGCATTCGGATCGTTTTGTCCTTATGTTGCTTTC 1671

Oy 121 ATGAGGGGGGAAAAAGCGGAAACCGCGAAGGTCAAAAGCTTTCGAGAGGTACTCTCA 180
Db 1672 ATGAGGGGGGAAAAAGCGGAAACCGCGAAGGTCAAAAGCTTTCGAGAGGTACTCTCA 1731

Oy 181 GCCTCATTTGAGCGCATTTATGATGAGGAGCGCTATGTTGAAAAAGTAACGAGCAAAA 240
Db 1732 GCCTCATTTGAGCGCATTTATGATGAGGAGCGCTATGTTGAAAAAGTAACGAGCAAAA 1791

Oy 241 GGGGAGAGCAGTATTTTCAATATGTTGAACCGATCTTTAGACCTCGGATTATGTACGA 300
Db 1792 GGGGAGAGCAGTATTTTCAATATGTTGAACCGATCTTTAGACCTCGGATTATGTACGA 1851

Oy 301 GGAACCTTGAACCCGGGTAAACCATCAAAAGATTTTAAACAAGCAGATTAAGAGATT 360
Db 1852 GGAACCTTGAACCCGGGTAAACCATCAAAAGATTTTAAACAAGCAGATTAAGAGATT 1911

Oy 361 CATCTGACAGCAATAGGAATCAGTGAAGCTCTTGAAGATATGAAATTTTCAAGCTTTC 420
Db 1912 CATCTGACAGCAATAGGAATCAGTGAAGCTCTTGAAGATATGAAATTTTCAAGCTTTC 1971

Oy 421 AACAGCGCAACACGCAATGATTAACGGCGTTCAGGGCGATGAAGATACGCTTGA 480
Db 1972 AACAGCGCAACACGCAATGATTAACGGCGTTCAGGGCGATGAAGATACGCTTGA 2031

Oy 481 GAATTTGGAAGCAAAACCTGATATCGTTGAGGGGAGTACAGTTAAGATGATCGGAA 540
Db 2032 GAATTTGGAAGCAAAACCTGATATCGTTGAGGGGAGTACAGTTAAGATGATCGGAA 2091

Oy 541 AAGAAAAATTTGCTACAGAAAGTCAACGGGGTAAAGATTTGCAACGCTTGCTTTACGAT 600
Db 2092 AAGAAAAATTTGCTACAGAAAGTCAACGGGGTAAAGATTTGCAACGCTTGCTTTACGAT 2151

Oy 601 GTGTCCGGGAAAGGTTTGGCGGCTTAAAAAGATACCGCGGCGTCTCCCGCAATCTT 660
Db 2152 GTGTCCGGGAAAGGTTTGGCGGCTTAAAAAGATACCGCGGCGTCTCCCGCAATCTT 2211

Oy 661 GAAATCTTCATCCCTATGATTTTCAAGAGGAAAAAATGCTGACATGTTGTTGTGAG 720
Db 2212 GAAATCTTCATCCCTATGATTTTCAAGAGGAAAAAATGCTGACATGTTGTTGTGAG 2271

Oy 721 TCACACTGGGGCAAGAGTATGACATGATCCAAACGACCGGCAAGCTTGCAAGA 780
Db 2272 TCACACTGGGGCAAGAGTATGACATGATCCAAACGACCGGCAAGCTTGCAAGA 2331

Oy 781 GCCATGCTGATGCGGAGCTGACATCATGTCGCCATCATCGCACGCTTTAGAACCG 840
Db 2332 GCCATGCTGATGCGGAGCTGACATCATGTCGCCATCATCGCACGCTTTAGAACCG 2391

Oy 841 ATTGAAGATATTAAGGAACCTCTATTTTCTACAGCTTCGCAACTTTTGTCTTTACCAA 900
Db 2392 ATTGAAGATATTAAGGAACCTCTATTTTCTACAGCTTCGCAACTTTTGTCTTTACCAA 2451

Oy 901 GGCTGAGAGAAACAAGAGACAGTGCACGTGTTCAAGTATCACTGAAGAAAAATGGAACA 960

```

```

Db 2452 GGCTGACGAGAAACAAGAGACAGTGCACGTGTTCAAGTATCACTGAAGAAAAATGGAACA 2511
Oy 961 GGCGGCTTTGAAGTACACCGATGATATCCATGAACGACACCTGACCGTGAAGAAA 1020
Db 2512 GGAGCGCTTTGAAGTACACCGATGATATCCATGAACGACACCTGACCGTGAAGAAA 2571
Oy 1021 GACAGCCTTAAACGAAAAACCATATTTCGCGAAGCTGACGAAAGACTAAATTTGCTTGG 1080
Db 2572 GACAGCCTTAAACGAAAAACCATATTTCGCGAAGCTGACGAAAGACTAAATTTGCTTGG 2631
Oy 1081 AAGTAGAAGCGGAAAACTGACGTTGATATGATGATGTCACAAACTTAAATCTTAA 1140
Db 2632 AAGTAGAAGCGGAAAACTGACGTTGATATGATGATGTCACAAACTTAAATCTTAA 2691

RESULT 7
BSUB0019/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
ORGANISM
REFERENCE
AUTHORS

Bacillus subtilis subsp. subtilis str. 168
Bacillus subtilis subsp. subtilis str. 168
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
1 (bases 1 to 200690)
Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G.,
Azevedo, V., Berteiro, M.G., Bessières, P., Bolo, A., Borchert, S.,
Borries, R., Boursier, L., Brans, A., Braun, M., Brimell, S.C.,
Bron, S., Brouillet, S., Brusch, C.V., Caldwell, B., Capuano, V.,
Carter, N.M., Choi, S.K., Codani, J.J., Conner, I.F., Cummings, N.J.,
Daniel, R.A., Denizot, F., Devine, K.M., Dusterhoft, A., Ehrlich, S.D.,
Emmerson, P.T., Entian, K.D., Errington, J., Fabret, C., Ferrati, B.,
Fouger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A.,
Galleron, N., Ghim, S.Y., Glaeser, P., Goffeau, A., Golligly, E.J.,
Grandi, G., Guisepi, G., Guy, B.J., Hage, K., Halech, J., Harwood, C.R.,
Henaux, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.F.,
Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y.,
Klaer, Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P.,
Koningstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A.,
Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H.,
Maesuda, S., Maue, C., Medigue, C., Medina, N., Mellado, R.P.,
Mizuno, M., Moesli, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M.,
Ogawa, K., Ogiwara, A., Oudega, B., Park, S.H., Parro, V., Pohl, T.M.,
Portecelle, D., Porwollik, S., Prescott, A.M., Presecan, E., Pujic, P.,
Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M.,
Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadate, Y., Sato, T.,
Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J.,
Sekowka, A., Seror, S.J., Serro, P., Shin, B.S., Soldo, B.,
Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K.,
Takeuchi, M., Tamakoshi, A., Tanaka, T., Terpetre, P., Tognoni, A.,
Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, P., Vassart, A.,
Viari, A., Wandt, R., Wedler, E., Wedler, H., Welter, T.,
Winters, P., Wipac, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K.,
Yoshida, K., Yoshikawa, H.F., Zumbstein, E., Yoshikawa, H. and
Danchin, A.

The complete genome sequence of the gram-positive bacterium
Bacillus subtilis
Nature 390 (6657), 249-256 (1997)
9384377
2 (bases 1 to 200690)
Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.
Direct Submission
Submitted (27-JUN-2003) I. Moszer, A. Danchin, Institut Pasteur,
Genetique des Genomes Bacteriens, 28 rue du Docteur Roux, 75724
Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
68 89 48
On Jul 7, 2003 this sequence version replaced gi:2636029.
This entry contains data from release R16.1 of the Substifiet
database. Further data on gene annotation and detailed information

```


Query Match	98.9%; Score 1127.2; DB 1; Length 200630;
Best Local Similarity	99.3%; Pred. No. 9.3e-276;
Matches 1132; Conservative	0; Mismatches 8; Indels 0; Gaps 0;
Qy	1 ATGAAAAAGAGCTTGAAGTCTTAAAGCTGCTAAAGCTGACAAAACAGCAAAAAAG 60
Db	89010 ATGAAAAAGAGCTTGAAGTCTTAAAGCTGCTAAAGCTGACAAAACAGCAAAAAAG 88951
Qy	61 AAAACCAATAAGCAGTATTTATGCGCATCGATTCGATTCCTATGTTGCTTC 120
Db	88950 AAAACCAATAAGCAGTATTTATGCGCATCGATTCGATTCCTATGTTGCTTC 88891
Qy	121 ATGTGGCGGGAAAAAGCGGAAACCGCGAAGGTCAAAAGTATTCGACGACTCTCA 180
Db	88890 ATGTGGCGGGAAAAAGCGGAAACCGCGAAGGTCAAAAGTATTCGACGACTCTCA 88831
Qy	181 GCCTCATTTGTAGGGAGATTTATGATGGAGCGCTATGTTGAAAAAGTAAACGAGCAAAA 240
Db	88830 GCCTCATTTGTAGGGAGATTTATGATGGAGCGCTATGTTGAAAAAGTAAACGAGCAAAA 88771
Qy	241 GGGGAGACAGTATTTTCAATATGTTAGCGCATCTTAGAGCTCGGATTTATGTAGCA 300
Db	88770 GGGGAGACAGTATTTTCAATATGTTAGCGCATCTTAGAGCTCGGATTTATGTAGCA 88711
Qy	301 GGAACCTTTGAAAAACCGGTAACTATCAAAAGATTATTAACAAGCAGATTAAGAGATT 360
Db	88710 GGAACCTTTGAAAAACCGGTAACTATCAAAAGATTATTAACAAGCAGATTAAGAGATT 88651
Qy	361 CATCTGCAGCAGATTAAGAAATCACTGAAAGTCTTGAAAGATATGATTTCAAGCTTTC 420
Db	88650 CATCTGCAGCAGATTAAGAAATCACTGAAAGTCTTGAAAGATATGATTTCAAGCTTTC 88591
Qy	421 AACGCGGCAACAACGACGCAATGATTACGGCGTTTCAAGGCGATGAAGATACGCTTGA 480
Db	88590 AACGCGGCAACAACGACGCAATGATTACGGCGTTTCAAGGCGATGAAGATACGCTTGA 88531
Qy	481 GAATTTGCGAAGCAAAACCTTGATTCGTTGAGCGGAGTACAGCTTAAGTATGTCGAAA 540
Db	88530 GAATTTGCGAAGCAAAACCTTGATTCGTTGAGCGGAGTACAGCTTAAGTATGTCGAAA 88471
Qy	541 AAGAAAATTTGCTACCAAGAGTCAACGGGGTAAAGATTGCAACGCTTGCTTACCGAT 600
Db	88470 AAGAAAATTTGCTACCAAGAGTCAACGGGGTAAAGATTGCAACGCTTGCTTACCGAT 88411
Qy	601 GTGTCCGGGAAAAGTTTGGCGCTAAAAAGAAATACCGCGGGGCTGCTCCGCAATCT 660
Db	88410 GTGTCCGGGAAAAGTTTGGCGCTAAAAAGAAATACCGCGGGGCTGCTCCGCAATCT 88351
Qy	661 GAAATCTTCATCCCTATGATTTCAAGAGCGAAAAACATGCTGACATTTGTTGTGAG 720
Db	88350 GAAATCTTCATCCCTATGATTTCAAGAGCGAAAAACATGCTGACATTTGTTGTGAG 88291
Qy	721 TCACACTGGGGCCAAAGATATGACATGATCCAAAGACCGCGCAGCCGCAAGTTCGAA 780
Db	88290 TCACACTGGGGCCAAAGATATGACATGATCCAAAGACCGCGCAGCCGCAAGTTCGAA 88231
Qy	781 GCCATGTCGATGCGGGAGCTGACATCATCGTCGCGCATCATCGGCAAGCTTTAGAACG 840
Db	88230 GCCATGTCGATGCGGGAGCTGACATCATCGTCGCGCATCATCGGCAAGCTTTAGAACG 88171
Qy	841 ATTGAAGTATATTAACGGAACCTGCTATTTCTACAGCTTCGCGCACTTTGTCTTTGACCA 900
Db	88170 ATTGAAGTATATTAACGGAACCTGCTATTTCTACAGCTTCGCGCACTTTGTCTTTGACCA 88111
Qy	901 GGCTGACGAGAACAGAGACAGTGCATGTTCAAGTATCACTGAAAGAAAAATGAAACA 960
Db	88110 GGCTGACGAGAACAGAGACAGTGCATGTTCAAGTATCACTGAAAGAAAAATGAAACA 88051
Qy	961 GGCGGCTTTGAAGTACACCGGATGATATCCATGAAAGGACGCTGCACTGTGAAAAA 1020
Db	88050 GGCGGCTTTGAAGTACACCGGATGATATCCATGAAAGGACGCTGCACTGTGAAAAA 87991
.Qy	1021 GACAGCCTTTAAACAGAAAAACATATTGCGGAATGACGAAAGACTTAATTTCCGTTGG 1080

Db	87990 GACAGCCTTTAAACGAAACCATTAATTCGGAACGTGACGAAAGCTTAATTTGCTTGG 87931
Qy	1081 AAAGTAGAAGCGGAAACAGCTTGAATATGATATGATGATGACAACTAAATCTTAA 1140
Db	87930 AAAGTAGAAGCGGAAACAGCTTGAATATGATATGATGATGACAACTAAATCTTAA 87871
RESULT 8	
LOCUS	DQ086153 2775 bp DNA linear BCT 27-JUN-2005
DEFINITION	Bacillus subtilis strain ZJU-7 Pgab (pgab), pgac (pgac), and pgaa
ACCESSION	DQ086153
VERSION	DQ086153.1 GI:68138277
KEYWORDS	
SOURCE	Bacillus subtilis
ORGANISM	Bacillus subtilis
REFERENCE	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
AUTHORS	Shi, F., Xu, Z., and Cen, P.
TITLE	1 (bases 1 to 2775)
JOURNAL	Efficient production of poly(glutamic acid) by a new strain
REFERENCE	Bacillus subtilis ZJU-7
AUTHORS	2 (bases 1 to 2775)
TITLE	Unpublished
JOURNAL	Submitted (04-JUN-2005) Institute of Bioengineering, Department of
REFERENCE	Chemical Engineering and Bioengineering, Zhejiang University, No.38,
AUTHORS	Zheda Road, Hangzhou, Zhejiang 310027, P.R. China
TITLE	Location/Qualifiers
FEATURES	
source	1..2775
	/organism="Bacillus subtilis"
	/mol_type="Genomic DNA"
	/strain="ZJU-7"
	/db_xref="taxon:1423"
gene	1..1182
	/gene="pgab"
CDS	1..1182
	/function="biosynthesis of poly(glutamic acid)"
	/codon_start=1
	/transl_table=1
	/product="pgab"
	/protein_id="AAV85940.1"
	/db_xref="GI:68138278"
gene	1183..1632
	/gene="pgac"
CDS	1183..1632
	/function="biosynthesis of poly(glutamic acid)"
	/codon_start=1
	/transl_table=1
	/product="pgac"
	/protein_id="AAV85941.1"
	/db_xref="GI:68138279"
gene	1633..2775
	/function="MFGSDYIALILVLSLIFAEKTIYVAGLVVPGVLGVNQP
	VFILLVLSLIVIVKYGSKRMILYGRKRFAMLTGVLKIAFDPLPIVPEI
	AERFGIYIPGLANTIKQGLITITGSLTLLSGATFALMFYLLI"
gene	1633..2775
	/gene="pgaa"
CDS	1633..2775
	/function="biosynthesis of poly(glutamic acid)"
	/codon_start=1

```

/translate=1
/product="PgaA"
/protein_id="AAV85942.1"
/db_xref="GI:68138280"
/translation="MKKELSPHEKLLKTKOOKKKNKHVEIPIVFLNFAFMWAG
KAEPKVTYDDVLSASFGVDIMGRVVEKTEBOKGDSI.FOYVE.FFRASDYVAGN
FENPTTYKRYKQADKEHLOTKNESVVLIDMNPETVLNSANNHMDGVQGMOTLG
EPAKONLDIVAGKISLDAKKKISYOKYNGVTIATLGTIDVSGRPAKKNPTGTLPA
DPEIPIPMISEAKKHADIIVQSHWGQEDNDPDROROLAKMANSDAQADITVGHHPH
VLEPIEVNGTIVIFYSIGNFVFDQMTFRSALVOYHLKNGTGRFEVPTDIHEAT
PAVVKDSLKQGTIIRELTKDSNFAMKVEDKLFEDIDHSDLKSK"

```

ORIGIN

```

Query Match      98.5%; Score 1122.4; DB 1; Length 2775;
Best Local Similarity 99.0%; Pred. No. 1.3e-274;
Matches 1129; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1 ATGAAAAAGAACTGATGCTTTCATGAAAAAGCTGTAAGCTGACAAACAGCAAAAAAG 60
DB 1633 ATGAAAAAGAACTGATGCTTTCATGAAAAAGCTGTAAGCTGACAAACAGCAAAAAAG 1692
QY 61 AAAACCAATTAAGCAGTATTTATGCAATCCGATCGTTTGTCTTATGTTGCTTTC 120
DB 1633 AAAACCAATTAAGCAGTATTTATGCAATCCGATCGTTTGTCTTATGTTGCTTTC 1752
QY 121 ATGTGGCGGGGAAAAAGCGGAAACGCCGAAGGTCAAAACGTATTCTGACGACTCTTCA 180
DB 1753 ATGTGGCGGGGAAAAAGCGGAAACGCCGAAGGTCAAAACGTATTCTGACGACTCTTCA 1812
QY 181 GGCTCATTTGTAGGGGATTTATGANTGGGACGCTATGTTGAAAAAGTAAACGAGCAAAA 240
DB 1813 GGCTCATTTGTAGGGGATTTATGANTGGGACGCTATGTTGAAAAAGTAAACGAGCAAAA 1872
QY 241 GGGGCGACAGTATTTTCAATATGTTGAACCGATCTTTAGAGCCTCGGATTATGAGCA 300
DB 1873 GGGGCGACAGTATTTTCAATATGTTGAACCGATCTTTAGAGCCTCGGATTATGAGCA 1932
QY 301 GGAAACTTTGAAACCCTGTAACCTATCAAAAGATTTATTAACAGCAAGTAAAGATTT 360
DB 1933 GGAAACTTTGAAACCCTGTAACCTATCAAAAGATTTATTAACAGCAAGTAAAGATTT 1992
QY 361 CATTCGACAGCAATTAAGATATGAGTAAAGTCTTGAAGATTAATTCACGGTTTC 420
DB 1993 CATTCGACAGCAATTAAGATATGAGTAAAGTCTTGAAGATTAATTCACGGTTTC 2052
QY 421 AACAGCGCAACCAACAGCAATGATTAACGGCTTCAGGGCATGAAAGTACGTTGGA 480
DB 2053 AACAGCGCAACCAACAGCAATGATTAACGGCTTCAGGGCATGAAAGTACGTTGGA 2112
QY 481 GAATTTGCAAGCAAAACCTTGATATCGTTGAGCGGGATACAGCTTAAGTATGCGAAA 540
DB 2113 GAATTTGCAAGCAAAACCTTGATATCGTTGAGCGGGATACAGCTTAAGTATGCGAAA 2172
QY 541 AAGAAAAATTTCTGACCAAGAAAGTCAACGGGGTACGATTGCAACGCTTGCTTACCGAT 600
DB 2173 AAGAAAAATTTCTGACCAAGAAAGTCAACGGGGTACGATTGCAACGCTTGCTTACCGAT 2232
QY 601 GTGTCCGGGAAAGTTTCGGGCTTAAAGAAAGATATACCGGGGCTGTGCGCGAGATCTT 660
DB 2233 GTGTCCGGGAAAGTTTCGGGCTTAAAGAAAGATATACCGGGGCTGTGCGCGAGATCTT 2292
QY 661 GAAATCTTCACTCCCTATGATTTGAGAGCGAAAAAACATGCTGACATTTGTTGTGCGAG 720
DB 2293 GAAATCTTCACTCCCTATGATTTGAGAGCGAAAAAACATGCTGACATTTGTTGTGCGAG 2352
QY 721 TCACACTGGGGCCAAAGATATGACATATGCAACGACCGCCACGCGCTTGCAAGA 780
DB 2353 TCACACTGGGGCCAAAGATATGACATATGCAACGACCGCCACGCGCTTGCAAGA 2412
QY 781 GCCATGCTGATGGGAGGTGACATCATGCTGGCGCATCATCCGACGCTTTAGAACCG 840
DB 2413 GCCATGCTGATGGGAGGTGACATCATGCTGGCGCATCATCCGACGCTTTAGAACCG 2472

```

```

QY 841 ATTGAAGTATATTAACGAAACCGTCAATTTTTCACAGCCTCGGCAACTTTGTCTTGACCA 900
DB 2473 ATTGAAGTATATTAACGAAACCGTCAATTTTTCACAGTCTCGGCAACTTTGTCTTGACCA 2532
QY 901 GGCTGACGAGAACAGACAGACAGTGCATGTTCAATATCACTGGAAGAAAAATGGAACA 960
DB 2533 GGCTGACGAGAACAGACAGACAGTGCATGTTCAATATCACTGGAAGAAAAATGGAACA 2592
QY 961 GGCCGCTTTGAAGTACACCGATGATTCATGTAAGGACACCTGCACCTGTGAAAAAA 1020
DB 2593 GGCCGCTTTGAAGTACACCGATGATTCATGTAAGGACACCTGCACCTGTGAAAAAA 2652
QY 1021 GACAGCCTTAAACAGAAACCAATTAATTCGCAACTGACGAAAGACTTAATTTGCTTGG 1080
DB 2653 GACAGCCTTAAACAGAAACCAATTAATTCGCAACTGACGAAAGACTTAATTTGCTTGG 2712
QY 1081 AAAGTGAAGACGGAACCTGACGTTGATATGATCATAGTACAAACTTAATCTTAA 1140
DB 2713 AAAGTGAAGACGGAACCTGACGTTGATATGATCATAGTACAAACTTAATCTTAA 2772

RESULT 9
AE017333 36/c
WPCOMMENT
Sequence split into 43 fragments LOCUS AE017333 Accession AE017333
Fragment Name Begin End
AE017333-00 1 110000
AE017333-01 100001 210000
AE017333-02 200001 310000
AE017333-03 300001 410000
AE017333-04 400001 510000
AE017333-05 500001 610000
AE017333-06 600001 710000
AE017333-07 700001 810000
AE017333-08 800001 910000
AE017333-09 900001 1010000
AE017333-10 1000001 1110000
AE017333-11 1100001 1210000
AE017333-12 1200001 1310000
AE017333-13 1300001 1410000
AE017333-14 1400001 1510000
AE017333-15 1500001 1610000
AE017333-16 1600001 1710000
AE017333-17 1700001 1810000
AE017333-18 1800001 1910000
AE017333-19 1900001 2010000
AE017333-20 2000001 2110000
AE017333-21 2100001 2210000
AE017333-22 2200001 2310000
AE017333-23 2300001 2410000
AE017333-24 2400001 2510000
AE017333-25 2500001 2610000
AE017333-26 2600001 2710000
AE017333-27 2700001 2810000
AE017333-28 2800001 2910000
AE017333-29 2900001 3010000
AE017333-30 3000001 3110000
AE017333-31 3100001 3210000
AE017333-32 3200001 3310000
AE017333-33 3300001 3410000
AE017333-34 3400001 3510000
AE017333-35 3500001 3610000
AE017333-36 3600001 3710000
AE017333-37 3700001 3810000
AE017333-38 3800001 3910000
AE017333-39 3900001 4010000
AE017333-40 4000001 4110000
AE017333-41 4100001 4210000
AE017333-42 4200001 4222645
Continuation (37 of 43) of AE017333 from base 3600001 (AE017333 Bacillus licheniformis D5
Query Match 47.4%; Score 540.2; DB 1; Length 110000;
Best Local Similarity 68.2%; Pred. No. 2.7e-126;

```

Matches 781; Conservative 0; Mismatches 358; Indels 6; Gaps 2;

QY 1 ATGAAAAAGAACTGAGCTTTTCATGAAAAAGCTGCTAAAGCTGACAAACAGCAAAAAAG 60
DB 65919 ATGAAAAAACAAGCACTTTTCAGAAAAAAGCTGGAAGTTGACGACAGAGAAAAAG 65860

QY 61 AAAACCAATAAGCAGTATTTATGCTCATTCGTTTGTCTTATGTTGCTTTC 120
DB 65859 AAAACCAATAAGCAGTATTTATGCTCATTCGTTTGTCTTATGTTTGTCTTTC 65800

QY 121 ATGTGGCGGGAAAAAGCGAAACGC--CGAAGCTCAAAACGTATTCTGACGAGTACTC 177
DB 65799 ACTTGGGTGGAGCGCCAAAACCTCTTCGCAAAATGCAAAAAAGAAAGATGCCAAGCTT 65740

QY 178 TCAGCCTATTGTGAGGAGTATTTATGATGAGCGCTATGTTGAAAAAGTAAACGAGACA 237
DB 65739 ACAGCTACTTTGTTGGGATATCATGATGAGAAACAGTAAAGAAAGTGAACAACTTG 65680

QY 238 AAAGGGGAGACAGTATTTTCAATATGTTGAAACCGATCTTTAGAGCTCGGATTATGTA 297
DB 65679 CACGGTTGGAAAGTGTCTTCAAAAATGTAAGCCGTACTTTATGTGTCAATTTTATC 65620

QY 298 GCAGGAACTTTGAAACCCGGTAACTATCAAAAAGAAATTATTAACAAGCAGATTAAG 357
DB 65619 ACAGGAACTTTGAAACCCGTAACTATCAAGAGCACTATCAAGAGCAGAAAAAGAC 65560

QY 358 ATTGATCTGACAGCAATTAAGGAATCAGTGAAGCTTTGAAAGATATGATTTCAAGCTT 417
DB 65559 ATCCATCTGCAACCAATCAAGAAATCAGTCAAAACATTGAAACCTGAACTTCAGCGTA 65500

QY 418 CTCAACAGCGCCAAACAACGACGATTAAGGATTAAGGCGGTTACGAGGATGAAGTACGCTT 477
DB 65499 CTGAATTTTGGCAACACCAATGCGATGACGCGGAAAGACGTTTGAAGATACGCTC 65440

QY 478 GGAGAAATTTGCGAAGCAAAACCTTGATATCTTTGAGCGGGATACAGCTTAAAGTATCG 537
DB 65439 AATAAATTTTCAATAGAAATCTGAGCTTGTGCGAGCAAGAAATATCTTGAAACCGG 65380

QY 538 AAAAAGAAATTTGCTTACCAAGAAATCAACGGGGTAAAGATGCAACGCTTGCTTAC 597
DB 65379 AAACGACGATCTTATCAAGAAATGAAACGGCGTAAATTTGCAACGCTGCTTTTACA 65320

QY 598 GATGTGTCGGGAAAGTTTCGCGCTAAAGAAATACGCGCGGCTGCTGCGCCAGAT 657
DB 65319 GACGTCTACAAAGAACTTTACACCAAAAAGAAAGAGGGGAGTGTCTGCGCTCAGT 65260

QY 658 CCGTGAATCTTCATCTCTATGATTTCAAGAGCGAAAAAATCATGTGACATGTTGTTGTG 717
DB 65259 CCGAAATCTTTATCTCAATGATTTGCGAAGCATGCAAAAAAGCGAATCTTGTCTTGTG 65200

QY 718 CAGTCACTGGGGGCAAGAGATGACAAATGATCCAAACGACCGCAGCGCAAGTTGCA 777
DB 65199 CATGTGCACTGGGCAAGAAATATGACAAATGAAACGAAACGAAAGAGATCTGACC 65140

QY 778 AGAGCCATGTCGATGCGGAGCTGACATCATGTCGCGCATCTCCGACGCTTTAA 837
DB 65139 AAGGCGATTCGAGATGCGGAGCAGATGTCATCATGCGCGCTCATCTCCATGTTCTGAA 65080

QY 838 CCGATTTGAATATATACGGAACCGTCAATTTTCAACGCTTCGCGCACTTGTCTTTGAC 897
DB 65079 CCGATTCGAAGTGTATACGGTACTGTGATTTTCTACGCTTCGCGCACTTGTATTTGAT 65020

QY 898 CAAAGCTTGACGAAACAAGACAGTGCATGCTTCAATCATCTGAGAAAAATGGA 957
DB 65019 CAGGCTGTGTCAAGAAACGCGGACAGCGGCTTGTACAAATCAATTTATGATGACGCG 64960

QY 958 ACAGGCGCTTTGAAGTGCACCGATCGATTCATTCATGGAAGCAGACTGCACT--GTG 1014
DB 64959 AAAGGCGCTTTGAGATACGCTCTCAACATTCGCGAAGCAACCGCAGCGCTTTAGGC 64900

QY 1015 AAAAAAGACAGCTTTAAACAGAAAAACATATTTGCGCACTGACAAAAAGACTTAATTTTC 1074
DB 64899 AAGACGACTTCTTAAAGGAAAGCGATCTTCCGTCAATTTGACAAAAAGAAACAACTTC 64840

QY 1075 GCTTGAAGTGAAGACGGAACCTGACGTTTGATATGATCATAGTACAAACTAAAA 1134
DB 64839 GACTGAAAGAAAGAAACGAAATTTAAGTTGAAGTCAATCATGCGGACAACTGAAA 64780

QY 1135 TCTAA 1139
DB 64779 AATAA 64775

RESULT 10
CP000002_36/c
WPCOMMENT
Sequence split into 43 fragments LOCUS CP000002 Accession CP000002

Fragment Name	Begin	End
CP000002_00	1	110000
CP000002_01	100001	210000
CP000002_02	200001	310000
CP000002_03	300001	410000
CP000002_04	400001	510000
CP000002_05	500001	610000
CP000002_06	600001	710000
CP000002_07	700001	810000
CP000002_08	800001	910000
CP000002_09	900001	1010000
CP000002_10	1000001	1110000
CP000002_11	1100001	1210000
CP000002_12	1200001	1310000
CP000002_13	1300001	1410000
CP000002_14	1400001	1510000
CP000002_15	1500001	1610000
CP000002_16	1600001	1710000
CP000002_17	1700001	1810000
CP000002_18	1800001	1910000
CP000002_19	1900001	2010000
CP000002_20	2000001	2110000
CP000002_21	2100001	2210000
CP000002_22	2200001	2310000
CP000002_23	2300001	2410000
CP000002_24	2400001	2510000
CP000002_25	2500001	2610000
CP000002_26	2600001	2710000
CP000002_27	2700001	2810000
CP000002_28	2800001	2910000
CP000002_29	2900001	3010000
CP000002_30	3000001	3110000
CP000002_31	3100001	3210000
CP000002_32	3200001	3310000
CP000002_33	3300001	3410000
CP000002_34	3400001	3510000
CP000002_35	3500001	3610000
CP000002_36	3600001	3710000
CP000002_37	3700001	3810000
CP000002_38	3800001	3910000
CP000002_39	3900001	4010000
CP000002_40	4000001	4110000
CP000002_41	4100001	4210000
CP000002_42	4200001	4222334

Continuation (37 of 43) of CP000002 from base 3600001 (CP000002 Bacillus licheniformis AT

Query Match 47.4%; Score 540.2; DB 1; Length 110000;
Best Local Similarity 68.2%; Pred. No. 2.7e-126;
Matches 781; Conservative 0; Mismatches 358; Indels 6; Gaps 2;

QY 1 ATGAAAAAGAACTGAGCTTTTCATGAAAAAGCTGCTAAAGCTGACAAACAGCAAAAAAG 60
DB 65759 ATGAAAAAACAAGCACTTTTCAGAAAAAAGCTGGAAGTTGACGAAAGCAGAAAAAG 65700

QY 61 AAAACCAATAAGCAGTATTTATGCTCATTCGTTTGTCTTATGTTGCTTTC 120
DB 65699 AAAACCAATAAGCAGTATTTATGCTCATTCGTTTGTCTTATGTTTGTCTTTC 65640

QY 121 ATGTGGCGGGAAAAAGCGAAACGC--CGAAGCTCAAAACGTATTCTGACGAGTACTC 177


```

Db 65639 ACTGGGTCGAGAGCGCCAAAATCTCTTCGCAATGACAAAAAGAAAGATCCAGCTT 65580
Qy 178 TCAGCCATTTTGTAGCGATATTATGATGGAGCCGATGTTGAAAAAGTAACGAGAGA 237
Db 65579 ACAGTACTTTTGTGGGATATCATGATGGAGAAAAGTAAGAAAAGTGAACAACTTG 65520
Qy 238 AAAGGGGAGACAGTATTTTCAATATGTTGAACCGATCTTTAGAGCTCGGATTATGA 297
Db 65519 CACGGTTCGAAAGTGTCTTCAAAAATGTGAAGCCGTCTTAAATGTGTCAATTTATC 65460
Qy 298 GCAGAAATCTTGAACCCGCTAATCAAAAGATTTAAACAGCAGATAAGAG 357
Db 65459 ACAGAAATCTTGAACCCGCTAATCAAAAGATTTAAACAGCAGATAAGAG 65400
Qy 358 ATTCACTGACAGACGATATAGGATCAGTGAAGTCTTGAAGATATGATTTCAAGCTT 417
Db 65339 ATCCATCTGCAACGATTCAGAAATCAGTCGAAACATTTGAAAAGCTGAACCTGACGCTA 65340
Qy 418 CTCAACAGCGCCCAACACGCAATGATTAAGCGCTTCAGGCGATGAAGATACGCTT 477
Db 65339 CTGAATTTTGGCCAAACACCATCGATGACTACGGGGAAGACGTTTGAAGATACGCTC 65280
Qy 478 GGAGAAATTTGCAACGAAACCTTGATATCGTTGAGGCGGATACAGCTTAAGTATGCG 537
Db 65279 AATMAATTTTCAAAATGAGATGTGAGCTTGTGAGAGCGAAATATCTTGAAGACGCG 65220
Qy 538 AAAAGAAATTTTCGACAGAAAGTCAACGCGGATTAACGATTTGCAACGCTTGCTTACC 597
Db 65219 AATACGACGATCTCTATCGAATGTGAACGCGATTAATTTGCAACGCTTGCTTACC 65160
Qy 598 GATGTGTCCGGGAAAGTTCGCGCTAAAGAAATACGCGCGGCTGTGCTCCGCAAT 657
Db 65159 GACGTCTACAAAGAACTTTACAGCCAAAGAAACAGAGCGGAGTGTGCGCTCACT 65100
Qy 658 CCTGAATTTTATCCCTTATGATTTCAAGAAAGGAAAAACATGCTGACATTTGTTGTTG 717
Db 65099 CGAAATATCTTTATTCATGATGTGGGAAAGCATGCAAAAAGCGATCTTGTCTTGTG 65040
Qy 718 CAGTCACTGGGGCAAGAGATGATGACATGATTCGCAACGCGGCGGCGGCGGATTTGA 777
Db 65039 CATGTGCACTGGGGCAAGAGATGATGACATGATTCGCAACGCGGCGGCGGATTTGA 64980
Qy 778 AGAGCATGTCTGATTCGCGGAGCTGACATCATGCTGCGCATCATTCGCAAGCTTTAGA 837
Db 64979 AAGGGATTTGAGATGCGGAGAGATGTCATCATGCGGCTCATCCCATGTTCTCGAA 64920
Qy 838 CCGATTTGAAGTATTAACGAAACCGTCAATTTTCTACAGCTCGGCAACTTTGCTTTGAC 897
Db 64919 CCGATCGAAGTGTATTAACGATCTGTGATTTTCTACAGCTCGGCAACTTTGATTTGAT 64860
Qy 898 CAAGGTGACGAGCAAGAGACAGTGCAGTGCAGTTCAGTATCACTGAAGAAATAGA 957
Db 64859 CAGGGTGTCTAAGAACAGGAGCAGCGGCTTGTCAATACATTTAATGAATGACGCG 64800
Qy 958 ACAGCGCTTTGAAGTGAACCGATCGATTCATGAAAGCGACACCTGACCT--GTG 1014
Db 64799 AAGGGCGCTTTGAGTAAAGCTCTCAACATTCGGAAGCAACGCGGCTTTAGGC 64740
Qy 1015 AAAAAAGACGCTTTAAACAGAAACCATTTATTCGGAAGTGAAGAAAGTAAATTTG 1074
Db 64739 AAGAGGACCTTTTAAACAGAAAGCATCTTCGTCATTTGAACAAAGGAAACAACTC 64680
Qy 1075 GCTTGAAGTGAAGAACGGAACGAGCTTGAATTTGATGATGATGATGATGATGATGAT 1134
Db 64679 GACTGGAAGAAAGAAAGGAAATTTAAGCTTTGAAGTGAATGCGGAGCAAGCTGAAA 64620
Qy 1135 TCTAA 1139
Db 64619 AATTA 64615

```

```

AX433330
LOCUS AX433330 696 bp DNA linear PAT 28-JUN-2002
DEFINITION Sequence 1745 from Patent WO0229113.
ACCESSION AX433330
VERSION AX433330.1 GI:21658134
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 Berka, R. and Clausen, I. G.
AUTHORS Methods for monitoring multiple gene expression
TITLE Patent: WO 0229113-A 1745 11-APR-2002;
JOURNAL Novozymes Biotech, Inc. (US) ; Novozymes A/S (DK)
FEATURES
source
1. 696
/organism="Bacillus licheniformis"
/mol_type="unassigned DNA"
/db_xref="taxon:1402"
ORIGIN
Query Match 29.2%; Score 332.4; DB 6; Length 696;
Best Local Similarity 68.4%; Pred. No. 1.8e-73;
Matches 475; Conservative 0; Mismatches 216; Indels 3; Gaps 1;
Qy 444 GGATTACGGGCTTCAGGGCATGAAAGATACGCTTGGAATTTGGGAGCAAAACCTTGA 503
Db 1 GACTACGCGGAAAGCGTTTGAAGATACGCTCAATATTTTCAATGAGATCTGGA 60
Qy 504 TATCGTTGAGCGGGATACAGCTTAAGTATGATCGGAAAGAAATTTGTCACAGAAAGT 563
Db 61 GCTTGTGAGAGAGAAATTAATCTTGAAGACGCGAAACGACGATCTCTATCAGAAATG 120
Qy 564 CAACGGGTTACGATTGCAACGCTTGCTTACCGATGTGTCCGGAAAGGTTTCCGGC 623
Db 121 GAACGCGTAAATTTGCAACGCTGCTTTTACAGACGCTTACACAAAGAACTTACAGC 180
Qy 624 TAAAGAAATACCGCGGCGCTGCTCCGCGAGATCTGTAATCTCATCCATGATTTG 683
Db 181 CAAAAAGAACAGAGCGGAGTGTCTCCCTCAGTCCGAAATCTTTATTCATGATTTGC 240
Qy 684 AGAAGCAAAAAACATGCTGACATTTGTTGTGACATCACACTGCGGGCAAGATATGA 743
Db 241 GGAAGATCGAAAAAGCGGATCTTGTCTTGTCCATGATGCACTGGGCAAGAAATATGA 300
Qy 744 CAATGATCGAAACGACCGGCAAGCGGCTTGAAGAGCATGTCTGATCGGAGCTGA 803
Db 301 CAATGACGAAACGACAGACAGAAAGATCTGGCAAGCGATTTGCAATGCGGAGCAGA 360
Qy 804 CATCATGCTGGGCAATCATCCGCAAGTCTTGAACCGATGGAAGTATTAACGGAACGT 863
Db 361 TGTCAATCATCGCGGCTCATCTCCCATGTTCTCGAACCGATGGAAGTATTAACGGAATGT 420
Qy 864 CATTTTCTACAGCTCGGCACTTTGTCTTTGACCAAGGCTGGAAGCAAGAGACAG 923
Db 421 GATTTTCTACAGCTCGGCACTTTGTATTTATTAAGGCTGTGTAAGAACAGGAGAG 480
Qy 924 TGCAGTGTTCAGTATCACTGAAGAAATGAAACGCGCTTTGAAGTGAACCGAT 983
Db 481 CCGGCTTTGACAAATCAATTTATGATGAACGGAAGGCGCTTTGAGTAAAGCTCTG 540
Qy 984 CGATATCATGAAGACGCTGCACT--GTGAAAAAGACAGCTTTAAACAGAAAC 1040
Db 541 CAACATTCGCGAAGCAACCGGCAAGCTTTAGCAAGACGCTTTTAAAGGAAAC 600
Qy 1041 CATTAATGCGAAGTGAAGAAAGCTTAATTTGCTTGAAGTGAAGAGCAAGAACT 1100
Db 601 GATCTTCGTCATTTGACAAAGAAACAACTTCGACTGAAAGAAAGAAAGAAATTT 660
Qy 1101 GACGTTGATTTGATCATATGACAAACTAAA 1134
Db 661 AACGTTGAAGTGCATGCGAGCAAGCTGAAA 694

```

RESULT 12
AB011191/c 94829 bp DNA circular BCT 14-JUN-2002
LOCUS Bacillus anthracis str. A2012 plasmid pXO2, complete sequence.
DEFINITION AE011191
ACCESSION AE011191
VERSION AE011191.1 GI:20520280
KEYWORDS
SOURCE Bacillus anthracis str. A2012
ORGANISM Bacillus anthracis str. A2012
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus; Bacillus cereus group.
REFERENCE 1 (bases 1 to 94829)
AUTHORS Read,T.D., Salzberg,S.L., Pop,M., Shumway,M., Umayam,L., Jiang,L., Holtzapple,E., Busch,J.D., Smith,K.L., Schupp,J.M., Solomon,D., Keim,P. and Frazer,C.M.
TITLE Comparative genome sequencing for discovery of novel polymorphisms in Bacillus anthracis
JOURNAL Science 296 (5575), 2028-2033 (2002)
PUBMED 12004073
2 (bases 1 to 94829)
Read,T.D., Salzberg,S.L., Pop,M., Shumway,M., Umayam,L., Jiang,L., Holtzapple,E., Busch,J.D., Smith,K.L., Schupp,J.M., Solomon,D., Keim,P. and Frazer,C.M.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2002) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
1. 94829
/organism="Bacillus anthracis str. A2012"
/mol_type="genomic DNA"
/strain="A2012"
/db_xref="taxon:191218"
/plasmid="pXO2"
/note="Florida strain"
complement (543. 1139)
/gene="BXB0001"
complement (543. 1139)
/gene="BXB0001"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein, (pXO2-03)"
/protein_id="AAM26163.1"
/db_xref="GI:20520281"
/translation="MTYVKLFQFRNGQIHVDIVDPQRFQPEYPRADVLRHAIYVLD EKREBGKQDDVSLKKEVDLQKQSVNTMQQMDVLKLMELIAQSGSNLSLEKLEQ EVKSEIASVTVKESKSGKPRPTVKREPIEENQEHQRIQREVKTSFVKKPLMITDDV QIWMKKGPQVVPRLQGTGYQEIEMEQIPHRKLE"
complement (1151. 1438)
/gene="BXB0002"
complement (1151. 1438)
/gene="BXB0002"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein, (pXO2-04)"
/protein_id="AAM26163.1"
/db_xref="GI:20520282"
/translation="MEIHVRNVDPYHLKEIDRKCEIGKLGRRYRYEYINMFEQH FNGEYSRKEDKPEDAVSNVITLDRQSDKLEIYIDATHELVAAMIKLNE"
complement (1652. 1768)
/gene="BXB0003"
complement (1652. 1768)
/gene="BXB0003"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAM26164.1"
/db_xref="GI:20520283"
/translation="MKRKKRYKVFRRIQSAVYRHIGNAYEVLKNAYKATHTI"
complement (1746. 2594)
gene

CDS
/gene="BXB0004"
complement (1746. 2594)
/gene="BXB0004"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein, (pXO2-05)"
/protein_id="AAM26165.1"
/db_xref="GI:20520284"
/translation="MSIKFPMNKEGRPATRKAYLFTLGSFVTFPVLCISPVSG ATRKFEWKIGESTQUSSTYKIAVAKKEINPDQTLRIDELRDNDSQLSMKRY ENRYIKQKDNVNTKYRASDNVYI VISENVPPEFQVSSVAVPEYTHPELQNDVLD KERSMKRYVLENKTLINRELKKSQDYERENYLAFSQALRKSEIKMEKSSAMQQL KIKNEOLTKMEYQTEGEKVTNKTINSNESTINNHQEKIDVLKEDIKMEKKIQQLD EKKKI"
complement (2591. 2866)
/gene="BXB0005"
complement (2591. 2866)
/gene="BXB0005"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein, (pXO2-06)"
/protein_id="AAM26166.1"
/db_xref="GI:20520285"
/translation="MLISELKPNNHDYAKESGYLLSLRKKKGVKQKFIPIWFDY NFGCEVWLVIVREYQSPVNGEKRYTNCKENIHQAQSVVAVKGBRMK"
complement (2955. 3563)
/gene="BXB0006"
complement (2955. 3563)
/gene="BXB0006"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein, (pXO2-07)"
/protein_id="AAM26167.1"
/db_xref="GI:20520286"
/translation="MNNLSKILSGALVLSFNGIMAYNTTKFEDKKQKRYRVTSK EKOIGDLKVLEOKDKKINSGAGENKEENSTLIDLNKRYEVANQFVHLYDVSVN KGERRNULKITDKKYVDIYAPNTDDGIDNPFNSHVYKAIYINSEBDVSKCTALLD IETYLEENKQYTTINSVKITLKEQEEIKVAVNYPYKR"
complement (3589. 4731)
/gene="BXB0007"
complement (3589. 4731)
/gene="BXB0007"
/note="identified by match to Pfam protein family HMM PF00877"
/codon_start=1
/transl_table=1
/product="NLP/P60 family protein, (pXO2-08)"
/protein_id="AAM26168.1"
/db_xref="GI:20520287"
/translation="MAITQETALDQYKKVRIKIVALLGSGTGVYAAATILLIYS MAIFGQSGSTGTPNGGISGTATVKNLPEPEVWRQAWMEQCAAGVELVPYVLAITM VESNGISEKLPIDIMQSSSESQMANNTISNPKDSIYGVNMLKGFDDAKMLGINDLLA IVOTYNGRNVYVHMLAANKKTHSIOTADYSLTVAVAGGNRNGTTIGSQPVAVAAN GGYRTVNGNFEVAVEMKYOLSPDAGTSGQIPDGSETPKMMDEVLYKNGNRYWVG GKSSGQFPDQSGILTYMAKYTAGLITPIISAATQVDFTENVDPKQAQPDGLVFFRRTYGG PNHVSHVGIYIDANTMTDNGSGVGHQFTSSVQGHYAGIRRVPR"
complement (4733. 6667)
/gene="BXB0008"
complement (4733. 6667)
/gene="BXB0008"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein, (pXO2-09)"
/protein_id="AAM26169.1"
/db_xref="GI:20520288"
/translation="MYLKGKKEQKRYTBEGYDDFISRVQPGGIRKNEKYIAGDCY VACHVYSLABEDIPPLWLTALMNNKDTISSVDATAKKEVVKQINSITELKQRTS ERSTDDDDHWELONTDPARSITQGEIVKLVKRIYIYDPLVLEDEKRIQDIKKE IAGNYKAQYVVFQKEWQTLFASYPDQILEYIKSGYPLPSKNIGFGIPFHHDLK

DRGIVIGTSTGCAFLIDPPESTGTRTSFSGFIRGKAGKSTLLKOLBGLVAKDC
 FIRGDKRARDYTVVVOOQCGKIDISGQDSEVEYTKMILNPLVPAATGTTMGANDKX
 GSFIQHSKVTMIRLFLNTEEDVIVQERKHLVFIEMGLVPOQSGDRPKVYGR
 PNHYILBDFPKYKLSNKLLEAGATPQRHDLALQVEMDITVYGDMEHNTLLKNF
 ENBOJLFDIDIGISKYDKSVFNCQLFTALQVEMDITVYGDMEHNTLLKNF
 YEMOJLFDIDIGISKYDKSVFNCQLFTALQVEMDITVYGDMEHNTLLKNF
 STMKAIFELTOYKIFLMDNSVLETLKSVLSESLRESERLPELKRGBAIVOVSTE
 TYNVMPDEKOLERFKGO"
 complement (6684. .7358)
 /gene="BxB009"
 /complement (6684. .7358)
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein, (px02-10)"
 /protein_id="AA026170.1"
 /db_xref="GI:20520289"
 /translation="MKMKVATAPAKKVRKROISEKSKGOSFISTIFASPKONTAS
 IVDIAPYKLEOKTGLVDHRRNLQVYKTKTDLISNODLXKRPANDTSLCXYH
 BPFKLSLSTETTEQOYVWRMALRFQGRMSQVSEKESHEWYQRYSLAENLR
 VLAVENKLKEAFIVVYGNKNETELIKVYKMKRYGGRQFNQNMKAKEVEELIKLO
 NNSSEM"
 complement (7416. .7730)
 /gene="BxB010"
 /complement (7416. .7730)
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein, (px02-11)"
 /protein_id="AA026171.1"
 /db_xref="GI:20520290"
 /translation="MEMMNPKNYTKQKILAFYIIGDAIILMLVLASMYKVPVVG
 GGRIMFYVLFGFGLFLCIKPHNSPTNNIFVILDMLKMDNKNVPIEVNTISSEK
 RK"
 complement (7733. .7936)
 /gene="BxB0011"
 /complement (7733. .7936)
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein, (px02-12)"
 /protein_id="AA026172.1"
 /db_xref="GI:20520291"
 /translation="MNNMIFAILMLGVALVLSIAIFPFQIPIFRKKKEPWWTV
 TNAEYIDRDPFKKKEQKDKKEH"
 complement (7938. .8192)
 /gene="BxB0012"
 /complement (7938. .8192)
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein, (px02-13)"
 /protein_id="AA026173.1"

Query Match 22.0%; Score 250.4; DB 1; Length 94829;
 Best Local Similarity 55.6%; Pred. No. 1.6e-52;
 Matches 576; Conservative 0; Mismatches 436; Indels 24; Gaps 4;

QY 1 ATGAAAAAGAACTGAGCTTTCATGAAAGCTGTAAGCTGACAAAGCAAGAAAAAG 60
 DB 55123 ATGAGAGCAAAATTCATTCATGAAAGAAAGTACTGATCTTTATTAAGAAACCAAGAAA 55064
 QY 61 AAAACCAATAGACGTAATTAATTCATTCGATCGTTTGTCTTAATGTTGCTTTC 120
 DB 55063 AAAAATCTCGTATAGCAATGTAATTAATTCATTCGATCGTTTGTCTTAATGTTGCTTTC 55004
 QY 121 ATGTGG---CGGAAAGAGCGAAAGCGCAAGGTCAAAAGTATTCGACGAGTACTC 177
 DB 55003 ACAATGGTACACGTACGAAGACGTAGACCAAGTAACATCGTGAAGCAAGAAATTTG 54944

QY 178 TCAGCCTATTGTAGGCGATATATGATGAGGACGCTATGTTGAAAAAGTAACGAGCAA 237
 DB 54943 ACATGACGATGCTGTTGGGACATTAATGATGAGGACGCTACGTAAGAAAGATTTGTAATGT 54884
 QY 238 AAAGGGGACAGACGTAATTTTCAATATGTTGAACCGATCTTTAGACCTCGGATTAATGA 297
 DB 54883 TACGCTACAGATTAATGTTTTCGTCATGTTTCGCCATTTTAAAAACCTCAGATTAACGA 54824
 QY 298 GCAGAAACTTTGAAAAACCC-----GGTAACCTATCAAAAGAAATTAATAACACACAGAT 351
 DB 54823 AGTGGAAATTTGAAACATCTGTTTGTAGAAAGTAATAAAAGATTAATAACACACAGAT 54764
 QY 352 AAAGATTTCAATCTGACAGCAAGTAAGAAATCAAGTAAAGTCTTGAAGATTAATGATTC 411
 DB 54763 AAGAAATTTCAATCTGACAGCAAGTAAGAAATCAAGTAAAGTCTTGAAGATTAATGATTC 54704
 QY 412 ACGTTCTCAACAGCGCCCAACCAACGCAATGATTAACGCTTCAAGGCGATGAAGAT 471
 DB 54703 ACAGATTAATTTGGCGAATACCATATGACGATTAATGTTGCTAAGGGAATTAAGAT 54644
 QY 472 ACGTTGAGAAATTTGCGAAGCAAAACCTGATATGTTGAGCGGAGTACAGCTTAAGT 531
 DB 54643 ACAATAAAGCCTTTAAAGAGCTGATCTGATATGTTGAGGCTGCTGTAATAATTTCAA 54584
 QY 532 GATGCGAAAAAGAAATTTTCGTAACAGAAAGTCAACGCGGTACGATTCACAGCTTGAC 591
 DB 54583 GATG---TAAATAAATTTGTTGATCAAAATGTAATTTGTTGAGGCTTCTACTTTGGA 54527
 QY 592 TTATCCGATGTTGCGGGAAGAGTTTTCGCGCTAAAGAAATATACCGCGGCTGCTGCC 651
 DB 54526 TTATCCGATGTTTGTGACAGAGCTATTCGACGAAAGAAACACAGGTTCTGTAAGT 54467
 QY 652 GCAGATCTGAAATCTTCAATCCCTATGATTTTCAAGAGCAAAAAAC-----AT 699
 DB 54466 ATGAACCCGATGTTATCTTAAGCAAAATTAAGTAAAGCAAGATCTTAAAGAGTAA 54407
 QY 700 GGTGACATTTGTTGTGTGAGTACACATGCGGGGCAAGATATGACATGATCAACAGAC 759
 DB 54406 GGTGATCTTGTGTGTGATTAATAGCACTGGGGGAAATAGATTAATTAACCGATCT 54347
 QY 760 CGCAGCGCCAGCTTCAAGAGCCATGTCGATGCGGAGTCAATCATCTGCGGCAT 819
 DB 54346 AGACAGAGAGCTTCAAGAGCCATGTCGATGCGGAGTCAATCATCTGCGGCAT 54287
 QY 820 CATCCGACGCTTTAGAACCGATGTAAGTATTAACGAAACCGTCAATTTTCTACAGCTTC 879
 DB 54286 CATCCGACGCTTTCAATCTTTGATGTTGATTAAGCAAGGATTAATCTTAATAGTTTA 54227
 QY 880 GCGAATTTGTTGTTGACCAAGGCTGACGAAAGCAAGACAGTCACTGGTTCAGTAT 939
 DB 54226 GGTAACTTTGTTGTTGACCAAGGATGACGAAAGCAAGACAGTCACTGGTTCAGTAT 54167
 QY 940 CACCTGAGAAATGAGAACGAGCGCTTTGAAAGTACACCATCGATATTCATGAAGCG 999
 DB 54166 CATTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 54107
 QY 1000 ACACCTGACCTGTGA 1015
 DB 54106 TCACCAAAACGATTA 54091

RESULT 13
 AE017335/c
 LOCUS 94830 bp DNA circular BCT 09-JUL-2004
 DEFINITION Bacillus anthracis str. 'Ames Ancestor' plasmid px02, complete
 ACCESSION AE017335
 VERSION AE017335.3 GI:50118566
 KEYWORDS
 SOURCE Bacillus anthracis str. 'Ames Ancestor'
 ORGANISM Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus

```
REFERENCE 1 (bases 1 to 94830)
AUTHORS Ravel,J., Raako,D.A., Shumway,M.F., Jiang,L., Cer,R.Z.,
          Federova,N.B., Wilson,M., Stanley,S., Decker,S., Read,T.D.,
          Salzberg,S. and Fraser,C.M.
TITLE    Bacillus anthracis comparative genomics
REFERENCE 2 (bases 1 to 94830)
AUTHORS Ravel,J., Raako,D.A., Shumway,M.F., Jiang,L., Cer,R.Z.,
          Federova,N.B., Salzberg,S. and Fraser,C.M.
TITLE    Direct Submission
JOURNAL  Submitted (17-MAY-2004) Microbial Genomics, The Institute for
          Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850,
          USA
REFERENCE 3 (bases 1 to 94830)
AUTHORS Ravel,J., Raako,D.A., Shumway,M.F., Jiang,L., Cer,R.Z.,
          Federova,N.B., Wilson,M., Stanley,S., Decker,S., Read,T.D.,
          Salzberg,S. and Fraser,C.M.
TITLE    Direct Submission
JOURNAL  Submitted (09-JUL-2004) Microbial Genomics, The Institute for
          Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850,
          USA
REMARK   Sequence update by submitter
COMMENT  On Jul 9, 2004 this sequence version replaced gi:47552342.
FEATURES
         source
             1..94830
             /organism="Bacillus anthracis str. 'Ames Ancestor'"
             /mol_type="genomic DNA"
             /strain="Ames Ancestor; A2084"
             /db_xref="taxon:261594"
             /plasmid="pXO2"
             /note="MLVA types G762 and A3_b"
             /locus_tag="GBAA_px02_0001"
             /complement(543..1139)
             /locus_tag="GBAA_px02_0001"
             /note="Identified by Glimmer2; putative"
             /codon_start=1
             /transl_table=11
             /product="hypothetical protein, (px02-03)"
             /protein_id="AAT28931.2"
             /db_xref="GI:47552343"
             /translation="MTYVKLFQFGNQIHVDIVDPFQRPYPEPREADVLRHATVLLD
             EKREKGGODVSLKKEVKDLKOSVNTKQOMDVLIKLMMELIAQSGNSLELEQ
             EVKSEIASVTVKSKSGKPPRTVKKREIPREINQOEHRIOREVKTSFVKKPLPMITDDV
             QIVAKKGRHVQVPRILCTGYOEIEMEDIPHRKKELE"
             /complement(1151..1438)
             /locus_tag="GBAA_px02_0002"
             /complement(1151..1438)
             /locus_tag="GBAA_px02_0002"
             /note="Identified by Glimmer2; putative"
             /codon_start=1
             /transl_table=11
             /product="hypothetical protein, (px02-04)"
             /protein_id="AAT28932.2"
             /db_xref="GI:47552344"
             /translation="MEIHVRVDPYHLKEIDKCKEIGKUGRPVYREYINMFEQH
             FNQYSRNKEDKFEBAVNSVITLDRQSDKIQEYIDATHELVAAVNIKLNKE"
             /complement(1652..1768)
             /locus_tag="GBAA_px02_0003"
             /complement(1652..1768)
             /locus_tag="GBAA_px02_0003"
             /note="Identified by Glimmer2; putative"
             /codon_start=1
             /transl_table=11
             /product="hypothetical protein"
             /protein_id="AAT28933.2"
             /db_xref="GI:47552345"
             /translation="MKRKKRYKXHFRIQSAVRIHQAVYEVLRKAYKAHTI"
             /complement(1746..2594)
             /locus_tag="GBAA_px02_0004"
             /complement(1746..2594)
             /locus_tag="GBAA_px02_0004"

gene
CDS
    /note="Identified by Glimmer2; putative"
    /codon_start=1
    /transl_table=11
    /product="conserved hypothetical protein, (px02-05)"
    /protein_id="AAT28934.2"
    /db_xref="GI:47552346"
    /translation="MSIKFWMNKGRRKPRATKRAVIFLTGSPYTFEFLCISPEBEG
    ATYFEEEMKIGEVQSLSTVIAVAKKEVDPQTLRI DYELRADNDQILSNMKRYV
    ENKTIKQKDNVNTKIVYRASDNVYIVISENVPEEPGVSSVVRKEIYHPELDNDVYL
    KEREMKRYVLENEKTLINRELKKSQDYEREBYLAFSQALRKETKEKDEKSSAMKQL
    KIKNEQGLTKMEVQTEBEKRYKTNTINSNSTINNHKEIDVLKEDIKMEKKIQGLD
    EKKKTI"
    /complement(2591..2866)
    /locus_tag="GBAA_px02_0005"
    /complement(2591..2866)
    /locus_tag="GBAA_px02_0005"
    /note="Identified by Glimmer2; putative"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein, (px02-06)"
    /protein_id="AAT28935.2"
    /db_xref="GI:47552347"
    /translation="MLSEIKPNHDVAKESKYLILSRKKGVKAKDFEIPITWFDY
    NFGKVEMLVIREYQPSVNGKEXYTNCKLENHQAQSVVAVKGRMK"
    /complement(2955..3563)
    /locus_tag="GBAA_px02_0006"
    /complement(2955..3563)
    /locus_tag="GBAA_px02_0006"
    /note="Identified by Glimmer2; putative"
    /codon_start=1
    /transl_table=11
    /product="conserved hypothetical protein, (px02-07)"
    /protein_id="AAT28936.2"
    /db_xref="GI:47552348"
    /translation="MNNLSKILSGALVLSIFNGCIWAVNTTKPEPKDKKRYRVTSRK
    EKQIGDLKVLQKQDKKINSAGAEENKKEGNSITDLQNKTRVANOQVYHAYLDVSNQ
    KGERRNLLKITDKKQVVDIVAPNTDDIGDNFQSHVAKAIIYNSBGDSKCTALLD
    IEYTIQELNKKQTTINSVKITTLEKQOEIKVVEYNBPYVR"
    /complement(3589..4731)
    /locus_tag="GBAA_px02_0007"
    /complement(3589..4731)
    /locus_tag="GBAA_px02_0007"
    /note="Identified by match to Pfam protein family HM
    PF00877"
    /codon_start=1
    /transl_table=11
    /product="NLP/P60 family protein, (px02-08)"
    /protein_id="AAT28937.2"
    /db_xref="GI:47552349"
    /translation="MAATQETALIDKRYKVRKIWIYLLGSGTGVIAAITLLIIVS
    MAIRGGSGTGPNGIGSTGATVKNLPEPWRHQMAVEQCAAGVELVYVLAIIIM
    VESNGISBKLPIDINQSESGQMANNTISNKDQSYVAVMLKGAFDAKMLGINDLLA
    IVQTFNGRNVVHMLANNKTHSIQTADYISLTVAVAGGRNGTTTGYQPAVAVYN
    GGYYIINGNFEVFAEMVQYLSFDGAGTSGQIPGSEETKRVMMDEVLKTNNGNVVWG
    KSSSGQFDCSGLTLYVAVYTAGITIPISATQYDFVPEVDPKADQPDLVFFRGTGG
    PNHVSIGIYIDANTMYDSNGSGVGHQFTSSVQCHYAGIRVRP"
    /complement(4733..6667)
    /locus_tag="GBAA_px02_0008"
    /complement(4733..6667)
    /locus_tag="GBAA_px02_0008"
    /note="Identified by Glimmer2; putative"
    /codon_start=1
    /transl_table=11
    /product="conserved hypothetical protein, (px02-09)"
    /protein_id="AAT28938.2"
    /db_xref="GI:47552350"
    /translation="MVILGKKEQKRYRTEGVDLPISRVPOGIGIKNEKXIAGDCY
    VACIVYSLAEDIPMLVLTALNNKQRTISSVDVATANKKEVQDINSITELKQMTS
    ERBSTDDDAHWEHLQNLTFPARSTTQGGIYVLRITYYDPLQLEKRIQDKRIS
    IAGQNYKQYVYFEQKEFMOTLPASVDOQIEYGVKSGYPLPSNIGRIPHHQIDIK
    DPGRIYAGQYSTGSAFTLDPFPSTGTSTSGFTPGMGAKSKTLLKQLEBGLVAKOC
    FIRGFDAQDYVTVVQOQGGKTIIDLQVQOSEEVETMIMPLEVFATKINTGAVDEK
    GSFIOHISKVTNMIRPLNTEFDDTVIOEPRKHLVAFYIEMGLVQKSDREYKVTGYP
```


CDS
/plasmid="PX02"
complement(19..477)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="PX02-01"
/protein_id="AAFI3606.1"
/db_xref="GI:6470152"
/translation="MRBGKGDVVLLKKEVKD.KOSVNMKQOMVLIKIMMELIAG
SGNSLSEKLEQEVKSEIASSVTKSEKPRPTKREPIENVEHQRIQEVKTTSE
VKRLPMITDDVQIVMKKQPHQVPRLGTYQEIEMEQIPEHRKELE"
complement(482..1018)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="PX02-02"
/protein_id="AAFI3607.1"
/db_xref="GI:6470153"
/translation="MGKSPSIIILTSQYTI.PNKKSFPTTVVDYCMRQKALLKDGHTLTNT
ELTSLAQNAIDSEPMRSGPHNQKVSSEKAEAKVLLTSPNSFTTDEKVV
SYMGRKALSKKKKTKTEEBEVSLLNHIIDLSPBEKKEGKGITAEITLGGVPSK
DKNVITSTDLNEIKDTIR"
complement(1037..1633)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="PX02-03"
/protein_id="AAFI3608.1"
/db_xref="GI:6470154"
/translation="MTYVKLFQFPGNOIHVDI.VODFQRPPEYREADVLRHATVLD
EKMEBGKQDDVSLKKEVKDLKOSVNMKQOMVLIKIMMELIAGSGNSLSEKLE
EVSKEIASSVTKSEKPRPTKREPIENVEHQRIQEVKTTSEVKRLPMITDDV
QIVMKKQPHQVPRLGTYQEIEMEQIPEHRKELE"
complement(1645..1932)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="PX02-04"
/protein_id="AAFI3609.1"
/db_xref="GI:6470155"
/translation="MEIHVNRVDPYHLKEIDKCKEIGKLGRRYYRMEYINMMPEQH
FNQGRSRKEDKPFQEAIVNSITLDRQDKQYVIDAHELVAAIMIKLNEE"
complement(2240..3088)
/note="similar to Dictyostelium discoideum
glutamine-asparagine rich protein encoded by Genbank
Accession Number U07817"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="PX02-05"
/protein_id="AAFI3610.1"
/db_xref="GI:6470156"
/translation="MSIKFMNKGKGRKPANTKRKAYLFTLGSFYTMFVLCISPVESG
ATYFEBMKIGEYQISSTVKIAYAKERYNDNQTIRIDYELRADNQISLNNKYV
ENKTIKQKDNVKTIVRASNDYIVVISENPBEFGVSSVAKPEYIHPBLQNDVLD
KERSMKYVLENEKILNELKSKDPFEREYLAQSQALKEIEKKEDSSAMKOL
KIKNEQLTKEMEYQTEGEEKVTKTINSESTINNHQKIDVLKEDIKKEKKIQLLD
EKKKTI"
complement(3085..3360)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="PX02-06"
/protein_id="AAFI3611.1"
/db_xref="GI:6470157"
/translation="MLSELKPNHDYAKEGYLLISLRKKKGVRRKDKFIEPIPIWFDY
NFGKEVEMIVREYQPSVNGKEKYNCKLENIHQAQSVVNVYGERMK"
complement(3449..4057)

CDS
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="PX02-07"
/protein_id="AAFI3612.1"
/db_xref="GI:6470158"
/translation="MNNLSKILSGALVLSFNGIWMAYNTTKSEFDKKKKRYVTSEK
EKOIGDLKVLQEQDKKINEGAGENKGNSTLIDONKYREVAQCFHAYLDSYQN
KGBRRNLKITDPRKQVIVAPMTDLDGDNFSPSHVKAIIYINSESDVSKCTALLD
IETIIELENKQTTINSVKITLEKQEEIKVYENPYPKR"
complement(4083..5225)
/note="similar to Enterococcus faecalis transposon TN916
ORF14 encoded by Genbank Accession Number AB014488"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="PX02-08"
/protein_id="AAFI3613.1"
/db_xref="GI:6470159"
/translation="MAATQETALIDYKVKRIRKIVRLGSGTGVIAAITLLIIVS
MAIFGQSGSTGNGISGTATYKNLPEYMRQAOVEBCAAGVBLPYVALIM
VESNGISEKPLDINQSESGMAMNTISNPKDSITYGVMHLKGFADAKMLGINDILA
IVQYFNGRNVVHMLAANKTHSIQIADYISLVIVAPAGNRKGTITGYSQPAVAYN
GGYRILINGNFYFAENVKQYLSFDGAGTSGQIPGSEETPKMMDEVLKYNMPPYVG
KNSSQGDFCSGLTYVAYKTAGITIPISAATQYDFVTEVDPKQAPDGLVFRTGYG
PNHSHVGIYIDANTWYDWSNGSGVGHQFTSSWQGHVAGIRRVPR"
complement(5227..7158)
/note="similar to Staphylococcus aureus transfer complex
protein trb encoded by Genbank Accession Number L11988"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="PX02-09"
/protein_id="AAFI3614.1"
/db_xref="GI:6470160"
/translation="MKLGKKEQQRRTGYDLDFISHPQGGIKENKXYIAQDCYV
ACLVHYSIADIDPLMLTALMNNKDTISSVDVATANKEEVVKDINRITELKRMTEB
RSTDRDADWELMLNTDFPARSITQGEIVKLVRLYIVDPVLEQLEKIGDKKEI
ACQNYKQVVPFKQKEWOTLPASYSDOIIVGVKSGYPLPSKNIGIGIPFHODLD
PRGIVLCQITSGAFILDPFPGTSTRISFSGFIIPGRGAKSTLLKQLEGVAKDDP
IRGFDKARDYTVVQOQ3SKIIDSQVDSBEVETGINLEVPATKINTGAVDEG
SFIOHISKVTNMRFLNTEFDDTVIQSFRLHVAFYIEWGLVPQKSDRYKVGYP
NHYIILEDPFYKYSNLLEAGATPORHRDEALKLQVEDMITYVGMFNHGTLLKNE
NEQIVFPDIDIGISKYDSVNCOLFALTLLIWSHALNGQMQLREKXLSIEDVY
FMVFLDCHVIVNSQNEFANRYVLEPERMRKFPAGIFPATQSPQEWLPPGSTDVS
TMKRIPELTQYKIFLANNDSVLETKSVLGSLSLSEFRLLPELRGBALVOVSSTET
YVWFDPDEKQLEKFKGQ"
complement(7178..7846)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="PX02-10"
/protein_id="AAFI3615.1"
/db_xref="GI:6470161"
/translation="MKVATAEAKVNRKQIESEKSGKOSFISTIPASPKNTDASIV
DIAYKLEKQGTGLVDRNNLOVLYLVKTTDLISNQDDLKRPMDLTSLCRVYHP
FKILSLVSTETTEQOYTKMKALRPGRSQSESEKKEHINVQISLAIBENRVL
WVENKLELAFPIVYVKNETELIKVNDMKRTGGROFNQNMKAKEVKLIFLQNM
NSEM"
complement(7910..8224)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="PX02-11"
/protein_id="AAFI3616.1"
/db_xref="GI:6470162"
/translation="MEMMRNPKNTKQIKALFPYIINDGAIILALMLVLSYMPKVPYVG
GFGRIIMPVLVGFITGLPLCIKPHNSPTNRNIFVLLDLKMDKNKYHPIEWNITSETEK
RK"
complement(8227..8430)

```

/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/translation="pX02-12"
/db_xref="GI:6470163"
/translation="MNWIFALMLGVALVLSITFRKQITFIKRRKKKEPVAVTV
TNAEYLDKDFWKGKQKQKKEHD"
complement(8432..8686)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/translation="pX02-13"
/protein_id="AAFI3618.1"
/db_xref="GI:6470164"
/translation="MKVIDIANRRRIYFPMKQOELRASILMTVAGFIIFAILVFOIS
BELGHLHYIYFAPLTYISLHLYSNKKARKIEKKQGY"
complement(8704..11562)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/translation="pX02-14"
/protein_id="AAFI3619.1"
/db_xref="GI:6470165"
/translation="MATMKIKEMFPOSKKALVIGILAVITLFSPTFPPIAYAPDWE
PGLMDVDEKVAFLONEMIOYADFLISQSIQGIKTLIYISALEGLIPDFS
PFLIKDNGIDNPASSIIGLFLYALFYIWMGLRTIIQHPKPRKSVGNILWIG
LIGALNELMADNMKSTDFSETTNTSKAKDGLANDIVQNTADLIYLSKTEPEPOS
KNEPEALISSTDSKPKNGMSKFIPLAQGDVATPEKIVMLTSQKDIPEKTEYIKI
TNDGKEIYKINDSVFNPFKDPGKYRYPWNPETIWMGLALGVALFLYFVFWM

```

Query Match 22.0%; Score 250.4; DB 1; Length 96231;
Best Local Similarity 55.6%; Pred. No. 1.6e-52;
Matches 576; Conservative 0; Mismatches 436; Indels 24; Gaps 4;

CDS

DB

QY 1 ATGAAAAAAGAACTGCTTTCATGAAAGCTGCTAAAGCTGACAAACGACGAAAAAG 60
55613 ATGAGACGAAATGACATTTCAAGAAAGTTACTGATCTTTAAGAAACGACGAAA 55554
QY 61 AAAACCAATGACGATATTATGCAATCCGATGCTTTTCTCTATGCTTC 120
55553 AAAAATCTCTGATATGCAATCGATTAATCTTATCGACATTAATTAATGCTGCG 55494
DB 121 ATCTGGG---CGGAAAGCGGAAAGCGGAGGTCAAAACGATTCGACGATCTC 177
55493 ACATGGGTACACGTAACGACGATGACCCAGTAACATGCTGAGAACGAAAAATTG 55434
QY 178 TCAGCTCATTTTGAAGCATATATGATGGAAGCTATGTTGAAAAAGTACGAGCAA 237
55433 ACATGACGATGTTGCTGATATGATGGAAGCTACGTAAGAAAGATGTTAATCGT 55374
DB 238 AAAGGGGACAGATATTTTCAATATGTTGAACCGATCTTTAGAGCTTGATATGTA 297
55373 TACGGTACAGATATGTTTTCGTCATGTTTCCCAATATTTAAAAAACTGAGATTACGTA 55314
QY 298 GCAGGAACTTTGAAAAACC-----GTTAACCTATCAAAAGAAATTAACAGAGAT 351
55313 AGTGGGAATTTGCAACATCTGTTTTGTGAAAGATTAAGAAATTAACAAAGAGAT 55254
DB 352 AAAGAGATTCATCTGACAGCAATTAAGCAATAGTGAAGCTTTGAAGATATGAAATTC 411
55253 AAGATATTTACTTAAGTCAAAAAGAAACAGTTAAGCAGTAAGAAAGAACCGGATTT 55194
QY 412 ACGGTTCTCAACAGCGCAACACGCAATGATTAAGCGCTTCAGGCGATGAAGAT 471
55193 ACGATATTAATTTGGGCAATTAACCATATGACGATTAATGCTTAAGGAACTAAAGAT 55134
QY 472 ACGGTTGAAGATTTGGGAAAGAAACCTTGATATCTTGAGAGGGGATACAGCTTAAGT 531
55133 ACATATAAGGCTTTAAAGAGCTGATCTGATATGATGATGCTGCTGTAAGAAATTTCAA 55074

QY 532 GATGCAAAAAAGAAATTTGCTACAGAAAGTCAACGGGCTAACGATTTGCAACGCTTGGC 591
55073 GATG---TAAATAATTTGTTGATATCAAAATGTAATGTTGTTAGGCTTGCTACTTGA 55017
QY 592 TTTCACGATGTCGCGGAAAGTTTTCGCGCTAATAAGAAATACCGCGGCTGCTGCC 651
55016 TTTCAGATGATTTGTCAGAGAGCTATTCACAAAGAAAGAAACAGTATTCGTTAAGT 54957
QY 652 GCAGATCTGAAATTTTATCCCTATGATTTTCAGAGGAAAGAAAC-----AT 699
54956 ATGAACCCAGATGATTTTCTTAAGCAATGTATGATGCAAGAGATCTTAAAAAGCTAAT 54857
QY 700 GCTGATTTGTTGTCAGATGATGATGCTGAGGCGCAAGATATGACATGATCAAAACGAC 759
54856 GCTGATTTGTCAGATGATGATGATGCTGAGGCGCAAGATATGACATGATCAAAACGAC 759
QY 760 CGCCAGCGCAGCTTCAAGAGCCATGCTGATGCGGAGCTGACATGATGCTGCGCAT 819
54836 AGACAGAAAGCTTCAAGAGCCATGCTGATGCGGAGCTGACATGATGCTGCGCAT 819
QY 820 CATCCGACGCTTTGAAACCATTTAATGATTAACGAAACCGTCAATTTCTACAGCCCT 879
54776 CATCCGACGCTTTGAAACCATTTAATGATTAACGAAACCGTCAATTTCTACAGCCCT 879
QY 880 GGCACTTTGCTTTGACCAAGGCTGACGAGAACAGACAGTCACTGCTGATAT 939
54716 GGTAATTTGCTTTGACCAAGGCTGACGAGAACAGACAGTCACTGCTGATAT 939
QY 940 CACTGAAAGAAATGAAACAGGCGCTTTGAAAGTGAACCGATGATTCATGAAGCG 999
54656 CATTACGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 999
DB 1000 ACACCTGACCTGTGA 1015
QY 54596 TCACCAAAACGATTA 54581

RESULT 15
BACCAPABC 3244 bp DNA linear BCT 26-APR-1993
LOCUS
DEFINITION
B. anthracis encapsulation protein genes (capA, capB, and capC),
complete cds.
M24150
M24150.1 GI:142630
VERSION
KEYWORDS
Bacillus anthracis
Bacillus anthracis
Bacillus anthracis
ORGANISM
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus; Bacillus
cereus group.
1 (bases 1 to 3244)
Makino, S., Uchida, I., Terakado, N., Sasaoka, C. and Yoshikawa, M.
Molecular characterization and protein analysis of the cap region,
which is essential for encapsulation in Bacillus anthracis
J. Bacteriol. 171 (2), 722-730 (1989)
2536679
JOURNAL
PUBMED
COMMENT
Original source text: B. anthracis (strain TE702, isolate pCap1)
DNA.
Draft entry and computer-readable sequence for [1] kindly provided
by I. Uchida, 01-MAY-1989.
Location/Qualifiers
1..3244
/organism="Bacillus anthracis"
/mol_type="genomic DNA"
/db_xref="taxon:1392"
282..1475
/note="44 Kd encapsulation protein CapB"
/codon_start=1
/translation="pX02-13"
/protein_id="AAFI3618.1"
/db_xref="GI:6470164"
/translation="MKVIDIANRRRIYFPMKQOELRASILMTVAGFIIFAILVFOIS
BELGHLHYIYFAPLTYISLHLYSNKKARKIEKKQGY"

CDS

DLAEALICECAVOPDYQIIFONKMIQANVIVNVLEDHMDVWGPTLDEVAEFTA
TIPYNGHLVITSESELDYFKEVEBERNTKIVVADNRSISEEFLKFDVMPDVAASIA
LVAEALGIDDETEAFRGLMNAHPDGMRIIRPADOSRPAFVFNPAANDSSRLTIM
ERVDPFGSNLAPVIMNCRPDRDRTEQFAPDULPYIKATIVAIGETTAPITSABE
KGDIPTOEYWNLEBMSWSEIWSRMKPYIKNIIVGVGNIHGAEPILDMTEBOIGKK
QAKVI"

CDS

1490..1939
/note="16 Kd encapsulation protein Capc"
/codon_start=1
/transl_table=1
/protein_id="AAA2287.1"
/db_xref="GI:142632"
/translation="MFSGDLYIALVLGVLTLIFERTGILPAGIIVPGYIALVFNOP
VEMVLVPLISITLVIVYGVSRFMLVGRKPAATLTGICLIFPCVPMPEI
PEFPGIGIVPLGLANTIQROGLPLTIGTITLLSGATPAINNIYLLP"
1951..3186
/note="46 Kd encapsulation protein Capa"
/codon_start=1
/transl_table=1
/protein_id="AAA2288.1"
/db_xref="GI:142633"
/translation="WRRLTPOEKLIIPIKTKKKNRYVAIVLPLIVIIAATWVO
RTEAVAPVHRENEKLIWTVWDIMGSHVEIVNRVGTVDYFHVSVYLLKSDVSG
NFEHPLVEDKKNYQKADKNLHLSAKEITVAVEAGFTVILANNHMTDYGATKXD
TIKAFKEADLDYVGAGENFKDKNIIVQNVANGVATLGTDAFVAGAIATKEQPGSL
SMNPDLIKOISKADPKKGNADLVVNTNGHEEDYDNKSPROEALAKAMDAGADLI
VGHHPHVLQSPDYVKQGIIFYSIGNFVDDQMTTRKDSALVOYHLRNDGTALIDVPL
NIQSGSPKPVASALDKRNVYROLTKDTSKGLMSKKDKLEIKLNHRIIVIKMKREK
QEHODKQENOVSEVETTT"

CDS

ORIGIN 748 bp upstream of HindIII site.
Query Match 21.9%; Score 250; DB 1; Length 3244;
Best Local Similarity 55.6%; Pred. No. 1.9e-52;
Matches 575; Conservative 0; Mismatches 435; Indels 24; Gaps 4;

QY 1 ATGAAAAAAGAACTGAGCTTTTCATGAAAAAGCTCTAAAGCTGACAAAACAGCAAAAAAG 60
DB 1951 ATGAGACGAAATTCACATTTCAAGAAAAAGTACGATCTTATTAAGAAAAACAAAGAA 2010
QY 61 AAAACCAATAGCAGTATTTTTCGCAATCCGATCGTTTGTCTTATGTCGCTTTC 120
DB 2011 AAAAATCCTCGTTATGTCATCGTATTAACCTTATCGCAGTTATTAATAGCTCG 2070
QY 121 ATGTGGG---GGGAAAAAGCGGAAAGCGGAAAGTCAAAAGTATTCGACGAGCTACTC 177
DB 2071 ACATGGGTACAACTGACGAAAGCACTAGCAGTAAACATCTGACAGACGAAAAATTG 2130
QY 178 TCAGCCTCATTTGTAGCGCATTTATGATGGAGCGCTATGTTGAAAAAGTAAACGAGCAA 237
DB 2131 ACGATGACGATGTTGGTACATTTATGATGGAGCGTCAAGTAAAGAGATTGTTAATCGT 2190
QY 238 AAAAGGGCAGACAGTATTTTCAATATGTTGAACGATCTTTAGAGCTTCGAGTTATGTA 297
DB 2191 TACGGTACAGATTATGTTTTCGTCATGTTTCCCATATTTAAAAAAGCTCAATTAACGTA 2250
QY 298 GGAGAAACTTGAAGAACCC-----GGTAACCTATCAAAAAGATTAATAACAGCAGAT 351
DB 2251 AGTGGGAATTTGAAACATCTCTGTTTGTGAAGATTAAGAAAGATTAATAACAGCAGAT 2310
QY 352 AAAAGATTTCATCTGACAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTTTC 411
DB 2311 AAGAAATATTCATTAAGTCAAAAAGAAAGACGTTAAAGCAGTAAAGAAAGCCGAGTTT 2370
QY 412 ACGGTTCTCAACAGCGCCAAACCAACGCAATGATTAACGCGTTCAAGGCGATGAAGAT 471
DB 2371 ACAGATTAAATTTGGCGAATTAACCATATGACGATTAATGATGCTAAAGGAACTAAAGAT 2430
QY 472 ACGCTTGGAATTTGGAGAGCAAAACCTGATATGTTGAGCGGGATACAGCTTAAGT 531
DB 2431 ACAATTAAGGCTTTAAAGAGCTGATCTTGACTATGTGGGTGCTGGTAAATTTTCAAA 2490
QY 532 GATGCGAAAAAGAAATTTCTGTAACGAAAGTCAACGGGCTAAACGATTGCAACGCTTGGC 591

DB 2491 GATG---TAAATAATTTGTGATCAAAATGTAATGATGTTAGGTTGCTACTCTTGA 2547
QY 592 TTATCCGATGTCGCGGAAAGGTTTTCGCGCTTAAAGAAATACGCCGGCGCTGCTGCC 651
DB 2548 TTTTCAATGATCATTTTGAAGAGAGCTATTTGCAACGAAAGAACACAGGTTGCTTAAGT 2607
QY 652 GCAGATCCTGAAATCTTCATCCCTATGATTTCAAGAACGAAAAAC-----AT 699
DB 2608 ATGAACCCAGATGATTAATTAAGCAAAATAGTAAGCAAAAGATCTTAAGAAAGTAAT 2667
QY 700 GCTACATTTGTTGTTGTCAGTACATCTGGGGCCAAAGTATGACATGATCCAAAGAC 759
DB 2668 GCTGATCTTTCGTAGTAATACGCACTGGGGGAAAGATACGATTAATTAACGAGTCT 2727
QY 760 CGCAGCGCAGCTTGCAGAGCCATGTCGATGTCGCGGAGCTGACATCTGTCGCCAT 819
DB 2728 AGACAGAAAGCTTGAAGAAAGCAATGTTGATGACAGGCGCAATATTTTGTGGACAC 2787
QY 820 CATCCGACAGCTCTTGAACCGATTGAAGTATTAACGAAACCGTCATTTTCTACAGCCTC 879
DB 2788 CATCCGATGATCTTCAATCTTTTGTATGTGATTAAGCAAGGATTAATCTTATAGTTTA 2847
QY 880 GCGAATCTTGTCTTTGACCAAGCTGACAGCAAGAACAGACACTGACATGCTCAATAT 939
DB 2848 GGTAACTTTGTTGTTGACCAAGGATGACAAAGAAACAAAGATAGTGCATTTGTGCAATAT 2907
QY 940 CACCTGAAGAAATGGAACAGCGCGCTTGAAGTGAACCGATCGATATTCATGAAGCG 999
DB 2908 CATTAGCTGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTA 2967
QY 1000 ACACCTGCACCTGT 1013
DB 2968 TCACCAAAACAGT 2981

Search completed: February 27, 2006, 11:11:11
Job time : 6570.47 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:36:56 ; Search time 6434.26 Seconds
(without alignments)
8289.569 Million cell updates/sec

Title: US-10-789-164-3

Perfect score: 1140
Sequence: 1 atgaaaaaagacgagctt.....gtgacaactaaatctaa 1140

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 41078325 seqs, 239354128 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: gb_esc1:*
2: gb_esc2:*
3: gb_esc3:*
4: gb_hcc:*
5: gb_esc4:*
6: gb_esc5:*
7: gb_esc6:*
8: gb_esc7:*
9: gb_gsa1:*
10: gb_gsa2:*
11: gb_gsa3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	DB ID	Description
C 1	45	3.9	1101 10	CNS0039G
C 2	43	3.8	471 7	CN774916
C 3	43	3.8	514 7	CN776311
C 4	43	3.8	790 10	CZ545842
C 5	43	3.8	1201 10	CNS0164Y
C 6	42.8	3.8	894 10	CNS10201
C 7	42.4	3.7	1101 10	CNS0181N
C 8	42.2	3.7	1020 1	AL564009
C 9	41.8	3.7	621 3	BI418173
C 10	41.6	3.6	342 1	AM740409
C 11	41.4	3.6	587 6	CA301908
C 12	41.4	3.6	694 7	CV151557
C 13	41.4	3.6	694 7	CV465225
C 14	41.2	3.6	534 5	B0844290
C 15	41.2	3.6	611 5	B0864323
C 16	41.2	3.6	643 5	B0854399
C 17	41.2	3.6	890 5	BX441261
C 18	41.2	3.6	1101 10	CNS0068P
C 19	41	3.6	920 10	CZ507711
C 20	41	3.6	930 10	CZ522718
C 21	40.6	3.6	738 10	CE355032
C 22	40.6	3.6	801 11	CR184582

C 23	40.4	3.5	724 10	CNS00780	AL066553 Drosophila
C 24	40.2	3.5	376 10	CE759802	CE759802 tigr-gss-
C 25	40.2	3.5	660 5	B0000205	B0000205 QG324D08.
C 26	40.2	3.5	669 5	B0857516	B0857516 QG37L17.Y
C 27	40.2	3.5	694 5	B0863273	B0863273 QG233H02.
C 28	40.2	3.5	703 5	B0861433	B0861433 QG18J09.
C 29	40.2	3.5	726 5	B0995701	B0995701 QG610L18.
C 30	40.2	3.5	744 5	B0864356	B0864356 QG26H09.
C 31	40	3.5	342 2	B3360474	B3360474 BR110646
C 32	40	3.5	360 1	AM740410	AM740410 BR110553
C 33	40	3.5	572 11	CR298214	CR298214 mrel-15L1
C 34	39.8	3.5	1191 3	BM471276	BM471276 AGENCOURT
C 35	39.6	3.5	270 5	BQ465530	BQ465530 HU03020F
C 36	39.6	3.5	542 5	BQ851926	BQ851926 QG16K16.
C 37	39.6	3.5	549 5	BQ851717	BQ851717 QG16B19.
C 38	39.6	3.5	562 5	B0865840	B0865840 QG6A13.Y
C 39	39.6	3.5	570 5	BQ992329	BQ992329 QG25F10.
C 40	39.6	3.5	581 5	BQ853621	BQ853621 QG21A03.
C 41	39.6	3.5	582 5	BQ852156	BQ852156 QG17E12.
C 42	39.6	3.5	591 5	BQ857564	BQ857564 QG17N18.Y
C 43	39.6	3.5	593 3	BI340328	BI340328 36558 MA
C 44	39.6	3.5	593 5	B0869460	B0869460 QG5D20.Y
C 45	39.6	3.5	599 5	BQ995224	BQ995224 QGPG14.Y

ALIGNMENTS

RESULT 1
LOCUS: CNS0039G 1101 bp DNA linear GSS 03-JUN-1999
DEFINITION: Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION: AL063921 GI:4941778
VERSION: AL063921
KEYWORDS: GSS.
SOURCE: Drosophila melanogaster (fruit fly)
ORGANISM: Drosophila melanogaster

REFERENCE: Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)

AUTHORS: Genoscope.
TITLES: Direct Submission
JOURNAL: Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage: BP 191 91006 EVRY cedex - FRANCE (E-mail: sequef@genoscope.cns.fr)

COMMENT

- Web : www.genoscope.cns.fr
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osogawa and Aaron Mammeter in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source
1. 1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR08K10"
/clone_11b="RPCI-98"
/note="end : TET3"

ORIGIN

Query Match	3.9%	Score 45	DB 10	Length 1101
Best Local Similarity	14.9%	Fred. No. 0.55		
Matches	62	Conservative 202	Mismatches 148	Indels 5
			Gaps	1
Oy	1	ATGAAAAAAGCAAGCAGCTTCATGATAAAGCTGCTAAAGCTGACAAAAACAGCAAAAAAG	60	
Db	1041	WWWMMWATWDTWMDKMMWMMATNATKTDPTWMTWRARWADNAGRGKAGKRDADATDAG	982	
Oy	61	AAAAACAATTAAGACGATATTTATTTGCGATTCGATCGTTTTGTCTTANGTTCGCTTTC	120	
Db	981	AGRRDGGKRDOKDOKDGGDDKGGKKKKAAKAKAKATWTKWDDWMDKMKMGAKDRK	922	
Oy	121	ATGTGCGGGGAAAAAGCGGAAACCCGGAAGTCAAAACGTATTTCGACGAGTACTGCA	180	
Db	921	ADDDGADKDDKDDGKXKADDDTGCTKDDDKDKKDDWDKAGTWGDATWMAATIDWWW	862	
Oy	181	GCTCATTTGTAGGCGATATTATGATGGACGCTATGTTGAAAAAGTAACGAGCAAAAA	240	
Db	861	GMADADWMTWMAADDMWADBRWDMAMKMDANAMGKRTADRRDWEGRAGKRGARRR	802	
Oy	241	GGGCGAG----ACGATATTTTTCATATGTTGAACCGATCTTTAGACCTCGGATATG	295	
Db	801	DRKADDRKADADRDADAATWTTTTRTDITDMDKMTDTWTRMAARTRDDBDDDD	742	
Oy	296	TAGCAGGAACTTTGAAAAACCCGGTAACTCTTCAAAAGATATATAACAGCAATAAG	355	
Db	741	RAGTARGRRTWRRTKRWRRDTRWDDADADTARDRRRRDDGADAKGKTGRKRRR	682	
Oy	356	AGATTCACTGTCAGACGAAATGAAGAACTGATGAAGAAGCTTGAAGATATGATTTCA	412	
Db	681	DRATWDRIDAMWADAMWTTTDTDTDDDKDRKRRKRRRTTARAAWMTWKA	625	
RESULT 2				
LOCUS	CN774916	471 bp	mRNA	linear
DEFINITION	tae75d10.y1 Hydra EST Dermatad I Hydra magnipapillata cDNA 5'			EST 20-MAY-2004
ACCESSION	CN774916			
VERSION	CN774916.1	GI:47545550		
KEYWORDS	EST.			
SOURCE	Hydra magnipapillata			
ORGANISM	Hydra magnipapillata			
REFERENCE	Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Anthomedusae; Hydridae; Hydra.			
AUTHORS	1 (bases 1 to 471)			
	Bode,H., Blumberg,B., Steele,R., Wigge,P., Gee,L., Nguyen,O., Martinez,D., Kibler,D., Hampson,S., Clifton,S., Page,D., Matra,M., Hillier,L., Martin,J., Wylie,T., Dante,M., Theisinger,B., Bowers,Y., Gibbons,M., Ritzer,E., Bennett,J., Ronko,I., Tsagarashvili,R., Maguire,L., Kennedy,S., Waterston,R. and Wilson,R.			
TITLE	Washu Hydra EST Project			
JOURNAL	Unpublished (2002)			
COMMENT	Contact: Hans Bode Washu Hydra EST Project Washington University School of Medicine 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu Library was constructed by Corina Guder / GATC Konstanz, Germany Library materials provided by Thomas Holstein / Molecular Cell Biology, TUD, Darmstadt DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Hans Bode (hrode@ucl.edu) Seq primer: -40UP. Location/Qualifiers 1..471 /organism="Hydra magnipapillata" /mol_type="mRNA" /strain="ef-1 mutant of Hydra magnipapillata" /db_xref="taxon:6085"			
FEATURES				
source				

Query Match	3.8%;	Score 43;	DB 7;	Length 471;
Best Local Similarity	49.8%;	Pred. No. 1.6;		
Matches 109;	Conservative 0;	Mismatches 110;	Indels 0;	Gaps 0;
Db	691	AAAAAATGTCGATCTGTTGTTGTGTCAGTCACATCTGGGGCCAAAGATATGACATGAT	750	
Qy	216	AAAAGTAAAGTGAATATTTATTTGTAACATTTTTCACCTGGGGAAACAGATATCTTGATTT	275	
Db	751	CCAAAGCAGCGCCAGCGCCAGCTTGGCAAGACCATGTCGTGATCGCGGAGCTGCATCATC	810	
Qy	276	CCAAAGAAACACAAAGAAATTTGGCAATTTATCTTAGTCATTTGGGTGTAAATTTATTC	335	
Db	811	GTGGCCATCATCTCCGCACTCTTGAAGACCATTTGAATATATATACGAAACCGCATTTTC	870	
Qy	336	ATTGTGTAGCCACCCACATGTAAATGCAAGACATGAATGGCTGAAACAATACCGTTGTCAT	395	
Db	871	TACAGCCCTGGCACTTGTGCTTGTGACCAAGAGCTGACG	909	
Qy	396	TATAGTTTGGGAATCTTGTTTTCATCCACATTTTCACG	434	
RESULT 3	CN776311	514 bp	mRNA	linear
LOCUS	CN776311			EST 20-MAY-2004
DEFINITION	caef79f10.y1 Hydra EST Darmstadt 1 Hydra magnipapillata cDNA 5' similar to TR:Q9RZD0 Q9RZD0 CONSERVED HYPOTHETICAL PROTEIN. ;, mRNA sequence.			
ACCESSION	CN776311			
VERSION	CN776311.1	GI:47546945		
KEYWORDS	EST.			
SOURCE	Hydra magnipapillata			
ORGANISM	Hydra magnipapillata			
REFERENCE	Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Anthomedusae; Hydridae; Hydra.			
AUTHORS	1 (bases 1 to 514) Bode,H., Blumberg,B., Steele,R., Wigge,P., Gee,L., Nguyen,O., Martinez,D., Kibler,D., Hampton,S., Clifton,S., Page,D., Marra,M., Hillier,L., Martin,J., Wylie,F., Dante,M., Theising,B., Bowers,Y., Glibbons,M., Rilter,E., Bennett,J., Ronko,I., Teagataishvili,R., Maguire,L., Kennedy,S., Waterston,R. and Wilson,R.			
TITLE	Maahu Hydra EST Project			
JOURNAL	Unpublished (2002)			
COMMENT	Other_ESTs: caef79f10.x1 Contact: Hans Bode Maahu Hydra EST Project Washington University School of Medicine 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: esb@wustl.edu Library was constructed by Corina Guder / GATC Konstanz, Germany Library materials provided by Thomas Holstein / Molecular Cell Biology, TUD, Darmstadt DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone Please contact: Hans Bode (hdbode@uci.edu) Seq primer: -40UP High quality sequence stop: 514. Location/Qualifiers 1..514 /organism="Hydra magnipapillata" /mol_type="mRNA" /strain="sf-1 mutant of Hydra magnipapillata" /db_xref="taxon:6085" /lab_host="Transformax ECl00 (Epicentre), T1 Phage resistant electrocompetent cells" /clone_lib="Hydra EST Darmstadt 1"			

ORIGIN

/note="Vector: pbluescript II SK (+); Site_1: NotI;
Site_2: EcoRI"

Query Match 3.8%; Score 43; DB 7; Length 514;
Best Local Similarity 49.8%; Pred. No. 1.6;
Matches 109; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

QY 691 AAAAAACATGCTGACATTTGTTGTGACGACACCTGGGCGCAAGATGATACATGAT 750
DB 227 AAAAGTAAAGTATGATATGTAACATTTTAACTGGGGAACAGAAATACCTTTCAT 286
QY 751 CCAACGACCGGACGCGGACGCTTGCAAGACCATCTGATCGGAGACTACATCATC 810
DB 287 CCAAAAGAAACCAAGAAATTTGGCAATTTTCTTAGCAATTTGGGTAAATTTAATC 346
QY 811 GTGGGACATCATCGGACGCTTTAGAACCGATTGATATATAGGAAACCGTCAATTTTC 870
DB 347 ATTTGATGACCCACCATGATATGCAAGACATGAATGCTGAACATACGCTTGTTCAT 406
QY 871 TACAGCTCTGGCACTTTGCTTTTGAACCAAGCTGAGC 909
DB 407 TATAGTTGGAAATCTGTTTTCATTCACATTTGACG 445

RESULT 4
CZ545842 790 bp DNA linear GSS 13-MAY-2005
LOCUS CZ545842
DEFINITION SRAA-aad64c12.g1 Strongyloides ratii whole genome shotgun library
(SRAAGS 004) Strongyloides ratii genomic, genomic survey sequence.
ACCESSION CZ545842
VERSION CZ545842.1 GI:64679122
KEYWORDS GSS.
SOURCE Strongyloides ratii
ORGANISM Strongyloides ratii
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Panagrolaimidae; Strongyloididae; Strongyloides.
1 (bases 1 to 790)
REFERENCE 1 Miteva,M., McCarter,J.P., Thompson,F., Viney,M., Page,D.,
Ritter,E., Martin,J., Wylie,T., Dante,M., Waterston,R.H.,
Clifton,S.W. and Wilson,R.
TITLE Genome Survey sequences from the rat parasitic nematode
Strongyloides ratii
JOURNAL Unpublished (2005)
COMMENT Washington University in St. Louis
Contact: Miteva M
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: nematode@watson.wustl.edu
Genomic DNA was provided by Fiona Thompson
(F.Thompson@bristol.ac.uk) and Mark Viney
(Mark.Viney@bristol.ac.uk) at the University of Bristol, Bristol,
UK.

FEATURES

source

Class: shotgun.

Location/Qualifiers

1..790

/organism="Strongyloides ratii"

/mol_type="genomic DNA"

/strain="Isotemale line ED321 heterogenic"

/db_xref="taxon:34506"

/dev_stage="infective larval stage (L3)"

/lab_host="GSI0"

/clone_11b="Strongyloides ratii whole genome shotgun
library (SRAAGS 004)"/note="Vector: POTW13; Site_1: BstXI; Site_2: BstXI;
Strongyloides ratii genomic DNA was randomly sheared,
end-repaired and size fractionated to enrich for 2-4 kb
fragments. Genomic DNA was provided by Fiona Thompson
(F.Thompson@bristol.ac.uk) and Mark Viney
(Mark.Viney@bristol.ac.uk) at the University of Bristol,
Bristol, UK. Sequencing by Washington University Genome

ORIGIN

Sequencing Center, St. Louis, MO."

Query Match 3.8%; Score 43; DB 10; Length 790;
Best Local Similarity 62.6%; Pred. No. 1.8;
Matches 67; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 815 GCCATCATCGGACGCTTTAGAACCGATTGAAGATATATACGGAACCGTCATTTCTACA 874
DB 7 GGCACCATCGGACATCATTTAGCCCTTGGAAATCTACAAAGACGTCATCTTCTACA 66
QY 875 GCGTCGCACTTTGCTTTTGAACCAAGGCTGACGAGAAAGAGAC 921
DB 67 GCGTCGCACTTTGCTTTGCGCTCGGCAATAGCAAGGAGAGGC 113

RESULT 5
CNS0164Y/c 1201 bp DNA linear GSS 26-JUL-1999
LOCUS CNS0164Y/c
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC
BACH15M04 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL106300
VERSION AL106300.1 GI:5621234
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1201)

REFERENCE 1 Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
Determination of this BAC end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.dbi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CBRH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11.

FEATURES

source

Location/Qualifiers

1..1201

/organism="Drosophila melanogaster"

/mol_type="genomic DNA"

/db_xref="taxon:7227"

/clone="BACH15M04"

/clone_11b="DrosBAC"

/plasmid="pBelobAC11"

/note="end : 17"

ORIGIN

Query Match 3.8%; Score 43; DB 10; Length 1201;
Best Local Similarity 35.6%; Pred. No. 2;
Matches 88; Conservative 46; Mismatches 113; Indels 0; Gaps 0;

QY 214 TATGTGAAGAAAGTAAAGGAGCAAAAGGAGACAGATTTTCAATATGTAACCG 273
DB 822 TTWATCATTTKATTAATAAAAAAAMAAAMAAATATTTTATTTTTHMMY 763
QY 274 ATCTTAGAGCTCGGATTTAGTATGACGAAACCTTGAACCCGTAACATCAAAAG 333
DB 762 CAGATTAATTTTDBADATTAAGCVRBARRRDRDDRDNDRWAMATATYCHWH 703
QY 334 AATTATTAACAGACATTAAGATTCATCTGACAGCAATTAAGAAATCAGTAAAGTC 393
DB 702 AATATATAGNCTGGCAACAAAMAAAGTCTTCTGCAAMAAAMAAATTCGKTAM 643
QY 394 TTGAAGATATGAATTTCAACGGTCTCAACAGGCCAACACGCAATGATTCGCGC 453

Db 642 ATAGACATTATTACGTATGATGATCACADAYBTGCGCMAATCAGSAMAAVACATWATKGT 583
QY 454 GTTCAGG 460
Db 582 WTTATG 576

RESULT 6
LOCUS CL510201/c 894 bp DNA linear GSS 01-APR-2004
DEFINITION SAIL_825_G02.v1 SAIL Collection Arabidopsis thaliana genomic clone
SAIL_825_G02.v1, genomic survey sequence.
CL510201-
CL510201.1 GI:4607521
GSS.
Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 894)
Sessions, A., Burke, E., Presting, G., Aux, G., McElver, J., Paton, D.,
Dietrich, B., Ho, P., Baccaden, J., Ko, C., Clarke, J. D., Cotton, D.,
Bullis, D., Snell, J., Miguel, T., Hutchinson, D., Kimmery, B.,
Mizel, T., Katsagiri, F., Glazebrook, J., Law, M. and Goff, S. A.
A high-throughput Arabidopsis reverse genetics system
Plant Cell 14 (12), 2985-2994 (2002)
12468722
COMMENT Contact: Sessions A
Applied Trait Genetics
Syngenta Biotechnology Inc.
3054 Cornwallis Rd., Research Triangle Park, NC 27709, USA
Email: allen.sessions@syngenta.com
ABRC Stock Number CS836839; T-DNA left border flanking sequences of
Syngenta Arabidopsis Insertion Library (SAIL) lines are available
through the Arabidopsis Biological Resource Center (ABRC).
Sequences represent a pool of amplified genomic regions and not
single contiguous sequences.
Class: TDNA tagged.
Location/Qualifiers
1. 894
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/ecotype="Columbia"
/db_xref="taxon:3702"
/clone_1ib="SAIL_825_G02.v1"
/note="T-DNA left border sequences were isolated using a
modified TAIL-PCR strategy"

ORIGIN
Query Match 3.8%; Score 42.8; DB 10; Length 894;
Best Local Similarity 56.3%; Pred. No. 2.2;
Matches 80; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Db 258 TCAATATGTGAACGATCTTAGAGCCTCGGATATTATGACAGAACTTGAATCC 317
QY 650 TAAATATTTTGAACCAAGACTTAAACAAAGCTATGTTATGAAAACCTTGAACCTT 591
Db 318 GGTAACTATCAAAAAGATTATTAACAGACAGATTAAGATTCATCTGCAGACGAATTA 377
QY 590 GAGAAACATCAAGATATGCAACACAAAAGTAAATGCAATATGAAACCGAA 531
Db 378 GGAATCAGTGAAGTCTTGAAG 399
QY 530 GGAGTAAGCAAAAGTGAAGAG 509

RESULT 7
LOCUS CNS0181N 1101 bp DNA linear GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC
BACN37P10 of DrosBAC library from Drosophila melanogaster (fruit

fly), genomic survey sequence.
ACCESSION AL108773
VERSION AL108773.1 GI:5629077
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
Genoscope.
REFERENCE Genoscope.
AUTHORS Direct Submission
TITLE Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11.
FEATURES
source Location/Qualifiers
1. 1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone_1ib="BACN37P10"
/clone_1ib="DrosBAC"
/plasmid="pBelobAC11"
/note="end : SP6"

ORIGIN
Query Match 3.7%; Score 42.4; DB 10; Length 1101;
Best Local Similarity 22.6%; Pred. No. 2.9;
Matches 83; Conservative 113; Mismatches 171; Indels 0; Gaps 0;

QY 86 CCATTGCGATGCTTTTGTCTTATGTTGCTTCATGTCGGCGGAAAGCGAAACGC 145
Db 579 CCCCCCAGATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 638
QY 146 CGAAGTCGAAACGATCTTGCAGCAGCTACTCTCAGCTCATTTGTAGCGATTTATGA 205
Db 639 GTGTTTGTGGGSGTTTGTGAGVGTKTKKGGGTTTTTTTMMMAAMADAR 698
QY 206 TGGAGCTATGTTGAAAAGTAACGAGCAAAAGGGGAGAGATTTTCAATATG 265
Db 699 GRGAVMGGSVAHKQKTKMMWRITKTSARAGRAAMGRSARMAAARTTHMMAAAAA 758
QY 266 TTGAACGATCTTTAGAGCCTCGGATATTATGACAGAACTTGAATCCCGTAACCT 325
Db 759 RAAMHARRAMTHHRRVARRRGRRRGGDRARRRRARRRRARRRRARRRG 818
QY 326 ATCAAAAGATTATTAACAGACAGATTAAGATTCATCTGCAGCAATTAAGAAATCAG 385
Db 819 RRMVRRMRRANNAABRGRARRRARRRARRRARRRARRRARRRARRRARRR 878
QY 386 TGAAGTCTTAAGAGATTAATTCATTCACGTTCTCAACAGCGCCACACACCAATCG 445
Db 879 AARARAAARRRRARRRARRRARRRARRRARRRARRRARRRARRRARRR 938
QY 446 ATTACGG 452
Db 939 RRRRAGG 945

RESULT 8
LOCUS ALS64009/c 1020 bp mRNA linear EST 05-APR-2004
DEFINITION ALS64009 Homo sapiens FETAL LIVER Homo sapiens cDNA clone
CSDDM001YE17 3-PRIME, mRNA sequence.
ACCESSION ALS64009

source

1. .342
/organism="Biomphalaria glabrata"
/mol_type="mRNA"
/strain="BS-90"
/db_xref="taxon:6526"
/clone="RBGIH55TR"
/sex="hermaphrodite"
/cell_type="hemocyte"
/lab_host="laboratory host"
/clone_lib="Biomphalaria glabrata (BS-90)-unexposed lambda zap library"
/note="vector: pbluescript SK-; site 1: EcoRI; site 2: XhoI; Total RNA was isolated from the hemocytes of unexposed Biomphalaria glabrata (BS-90) snails and first strand cDNA synthesized using an oligo-dT primer-linker (XhoI). Second strand synthesis was followed by the ligation of EcoRI adaptors. Following digestion with XhoI, the completed, directional cDNA was cloned into Uni-ZAP XR Phagemid vector by Stratagene."

ORIGIN

Query Match 3.6%; Score 41.6; DB 1; Length 342;
Best Local Similarity 50.0%; Pred. No. 3.7;
Matches 104; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 22 CATGAAAGCTGTAAAGCTGACAAACAGCAAAAGAAACCAATAGACGATTT 81
DB 2 CACGAGCATGATGATGAGAGGTAAAGCTCACGACTGACACCAATGATCTTT 61
QY 82 ATTGCATTCGATCGATCTTTTCTTCTTATGTTGCTTCATGTGGCGGAAAGCGGA 141
DB 62 GGTGCTATCATGCTCTATTTAGACATCATGATGTTTTCTGTACATCCTTATCATGCG 121
QY 142 ACGCCGAAGGTCAAAAGCTATTTTACGACGATCTCAAGCTCTATTGTAGCGCATTT 201
DB 122 GACTCGACCTGAGAAACATATTCAAGTCTTAACCTCAATTTCCAAATCAGAAAGAAAA 181
QY 202 ATGATGGAGCGCTATGTTGAAAGTAA 229
DB 182 ATCTAGACAGAAACATTTGAAGAACTAA 209

RESULT 11
CA301908/c 587 bp mRNA linear EST 01-NOV-2002
LOCUS caa10d03.x1 Hydra cDNA library Hydra magnipapillata cDNA 3' similar
DEFINITION to TR:Q9RZD0 Q9RZD0 CONSERVED HYPOTHETICAL PROTEIN.; mRNA
SEQUENCE.
ACCESSION CA301908
VERSION CA301908.1 GI:24464967
KEYWORDS EST.
SOURCE Hydra magnipapillata
ORGANISM Hydra magnipapillata
Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Anthomedusae;
Hydridae; Hydra.
REFERENCE 1 (bases 1 to 587)
Bode,H., Blumberg,B., Steele,R., Wigge,P., Gee,L., Nguyen,Q.,
Martinez,D., Kibler,D., Hampson,S., Clifton,S., Page,D., Marra,M.,
Hillier,L., Martin,J., Wylie,T., Dante,M., Theising,B., Bowers,Y.,
Gibbons,M., Ritter,E., Bennett,J., Ronko,I., Tsagarisvilli,R.,
Maguire,L., Kennedy,S., Waterston,R. and Wilson,R.
TITLE WASHU HYDRA EST PROJECT
JOURNAL Unpublished (2002)
COMMENT Contact: Hans Bode
Washu Hydra EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
Library was constructed by H. Bode and B. Blumberg DNA sequencing
by: Washington University Genome Sequencing Center For information
on obtaining a clone please contact: Hans Bode (hnbode@ucl.edu)

Seq primer: Primer name ambiguous
High quality sequence stop: 404.
Location/Qualifiers

1. .587
/organism="Hydra magnipapillata"
/mol_type="mRNA"
/strain="105"
/db_xref="taxon:6085"
/lab_host="DHSalpa"
/clone_lib="Hydra cDNA library"
/note="vector: pSPORT6; site 1: NotI; site 2: SalI;
libraries prepared by Phil Wigge, Plant Molecular and
Cellular Biology, Lab (Detlev Weigell), The Salk Institute
for Biological Science, 10010 North Torrey Pines Road, La
Jolla, CA 92037."

ORIGIN

Query Match 3.6%; Score 41.4; DB 6; Length 587;
Best Local Similarity 49.3%; Pred. No. 4.8;
Matches 108; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

QY 691 AAAAATGATGATGATTTGTTGTCAGTCACTGCGGCGCAAGTATGACATGAT 750
DB 554 AAAAGTAAAGTATATTTATTTATGATTTTACACTGCGGAAACAGATCTTGTAT 495
QY 751 CCAAGACCGCCGAGCCGCTTCAAGAGCCATGTCGAGCGGAGCTGACATCATC 810
DB 494 CCAAAAGAAACAAAGAAATTTGCAATTTATTTGTCATTTGGGTGTAATTTATC 435
QY 811 GTGCGCATCATTCGCGACGCTTTAGAAACGATGATATATTAACGAAACCGTATTTC 870
DB 434 ATGTGTAGCCACCCACATGATGCAAGACATGAGTGGCTGAAACAATAGCTTTTCAT 375
QY 871 TACAGCTCGGCACTTTGTTTGAACCAAGCGTGGACG 909
DB 374 TATAGTTTGGAAATCTTTTTCATTCACATTTACG 336

RESULT 12
CV151557/c 694 bp mRNA linear EST 08-SEP-2004
LOCUS ca171e03.x2 Hydra EST UCI 5 Alp Hydra magnipapillata cDNA 3'
DEFINITION similar to TR:Q9RZD0 Q9RZD0 CONSERVED HYPOTHETICAL PROTEIN.; mRNA
SEQUENCE.
ACCESSION CV151557
VERSION CV151557.1 GI:51945203
KEYWORDS EST.
SOURCE Hydra magnipapillata
ORGANISM Hydra magnipapillata
Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Anthomedusae;
Hydridae; Hydra.
REFERENCE 1 (bases 1 to 694)
Bode,H., Blumberg,B., Steele,R., Wigge,P., Gee,L., Nguyen,Q.,
Martinez,D., Kibler,D., Hampson,S., Clifton,S., Page,D., Marra,M.,
Hillier,L., Martin,J., Wylie,T., Dante,M., Theising,B., Bowers,Y.,
Gibbons,M., Ritter,E., Bennett,J., Ronko,I., Tsagarisvilli,R.,
Maguire,L., Kennedy,S., Waterston,R. and Wilson,R.
TITLE WASHU HYDRA EST PROJECT
JOURNAL Unpublished (2002)
COMMENT Other ESTs: ca171e03.y2
Contact: Hans Bode
Washu Hydra EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
Library was constructed by Dirk Lindgens, Univ. of Calif., Irvine
Library materials provided by Hans Bode & Dirk Lindgens, Univ. of
Calif., Irvine DNA sequencing by: Washington University Genome
Sequencing Center For information on obtaining a clone please
contact: Hans Bode (hnbode@ucl.edu)
Seq primer: -40UP from Gibco

TITLE
JOURNAL
COMMENT
 Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J.,
 Ellison, P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y.,
 Lai, Z., Church, S., Jackson, L., and Bradford, K.
 Lettuce and Sunflower ESTs from the Composite Genome Project
 http://compgenome.ucdavis.edu/
 Unpublished (2002)
 Contact: Alexander Kozik [R.W.Michelmores]
 Department of Vegetable Crops, R.W.Michelmores Lab
 University of California at Davis (UCD)
 Armadson Hall, UCD, Davis, CA 95616, USA
 Tel: 1-(530)-742-1742
 Fax: 1-(530)-752-9659
 Email: akozik@ucdavis.org [michelmores@vegmail.ucdavis.edu]
 belongs to contig OG_CA_Contig6559, see http://cgpb.ucdavis.edu/
 for details.
 Plate: OG13 row: H column: 12.
 Location/Qualifiers
 1..534
 /organism="Lactuca sativa"
 /mol_type="mRNA"
 /cullivar="Salinas"
 /db_xref="taxon:4236"
 /clone="OG13H12"
 /lab_host="E.coli"
 /clone_lib="OG ABCDI lettuce salinas"
 /note="Vector: pBRCDNASFIAB; The library was constructed
 from 10 different sources of RNA from a single genotype.
 Separate cDNAs were generated using primers that
 incorporated unique 5' and 3' tags to distinguish each
 source of RNA. cDNAs were then pooled, size-fractionated,
 directionally cloned into a custom medium-copy vector and
 transformations made with four size classes to minimize
 size bias. Details of each source of RNA and library
 construction can be obtained at http://cgpb.ucdavis.edu/
 TAG_SEQ=Not found"

ORIGIN
 Query Match 3.6%; Score 41.2; DB 5; Length 534;
 Best Local Similarity 50.5%; Pred. No. 5.3; Mismatches 0; Gaps 0;
 Matches 100; Conservative 0; Indels 0; Gaps 0;
 448 TACGGCGTTACAGGCGATGAAGATACGCTTGAGAAATTTGCGAACAAACCTTGATATC 507
 309 TCCGCTTCGAGATCTGAAGCATCATGTGAAGCATCGGATCGCTTCCTGATTTTC 250
 508 GTTGAGCGGGATACGCTTAAGTATGCGAAAAAATTTCTTACCAAAAGTCAAC 567
 249 GTCGAAGAGACGACGGGTGAGGGGCGGCGGAACCGCTTCGTTAAGTGAACCGCTTC 190
 568 GGGGTAAAGATTGCAACGCTTGATCCGATGTGTCGGGAAAGTTTCCGGGCTAAA 627
 189 GGTGAGCAACGTAAACGGGGCGGTGAACCGATGAGCTTTGAGACGGTGTGCGTTCCAT 130
 628 AAGAATACGCCGGGCGGTG 645
 129 GAACCTCACTCATGTGAG 112

RESULT 15
 BO864323/c 611 bp mRNA linear EST 14-AUG-2002
 LOCUS OGCG26J20.yg.ab1 OG ABCDI lettuce salinas Lactuca sativa cDNA clone
 DEFINITION OGCG26J20.yg.ab1 OG ABCDI lettuce salinas Lactuca sativa cDNA clone
 ACCESSION BO864323
 VERSION BO864323.1 GI:22249869
 KEYWORDS EST.
 SOURCE Lactuca sativa
 ORGANISM Lactuca sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
 asterids; Campanulids; Asterales; Asterales; Cichorioideae;
 Cichorioideae; Lactuca.
 .REFERENCE
 1 (bases 1 to 611)

AUTHORS
 Kozik, A., Michelmores, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,
 Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J.,
 Ellison, P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y.,
 Lai, Z., Church, S., Jackson, L., and Bradford, K.
 Lettuce and Sunflower ESTs from the Composite Genome Project
 http://compgenome.ucdavis.edu/
 Unpublished (2002)
 Contact: Alexander Kozik [R.W.Michelmores]
 Department of Vegetable Crops, R.W.Michelmores Lab
 University of California at Davis (UCD)
 Armadson Hall, UCD, Davis, CA 95616, USA
 Tel: 1-(530)-742-1742
 Fax: 1-(530)-752-9659
 Email: akozik@ucdavis.org [michelmores@vegmail.ucdavis.edu]
 belongs to contig OG_CA_Contig6559, see http://cgpb.ucdavis.edu/
 for details.
 Plate: OGCG26 row: J column: 20.
 Location/Qualifiers
 1..611
 /organism="Lactuca sativa"
 /mol_type="mRNA"
 /cullivar="Salinas"
 /db_xref="taxon:4236"
 /clone="OGCG26J20"
 /lab_host="E.coli"
 /clone_lib="OG ABCDI lettuce salinas"
 /note="Vector: pBRCDNASFIAB; The library was constructed
 from 10 different sources of RNA from a single genotype.
 Separate cDNAs were generated using primers that
 incorporated unique 5' and 3' tags to distinguish each
 source of RNA. cDNAs were then pooled, size-fractionated,
 directionally cloned into a custom medium-copy vector and
 transformations made with four size classes to minimize
 size bias. Details of each source of RNA and library
 construction can be obtained at http://cgpb.ucdavis.edu/
 TAG_TISSUE=chemical induction
 TAG_LIB=OG ABCDI lettuce salinas
 TAG_SEQ=GTGACCGCGG"

ORIGIN
 Query Match 3.6%; Score 41.2; DB 5; Length 611;
 Best Local Similarity 50.5%; Pred. No. 5.5; Mismatches 98; Indels 0; Gaps 0;
 Matches 100; Conservative 0; Indels 0; Gaps 0;
 448 TACGGCGTTACAGGCGATGAAGATACGCTTGAGAAATTTGCGAACAAACCTTGATATC 507
 332 TCCGCTTCGAGATCTGAAGCATCATGTGAAGCATCGGATCGCTTCCTGATTTTC 273
 508 GTTGAGCGGGATACAGCTTAAGTATGATCGAAAAAATTTGTCACGAAAGTCAAC 567
 272 GTCGAAGAGACGACGGGTGAGGGGCGGCGGAACCGCTTCGTTAAGTGAACCGCTTC 213
 568 GGGGTAAAGATTGCAACGCTTGATCCGATGTGTCGGGAAAGTTTCCGGGCTAAA 627
 212 GGTGAGCAACGTAAACGGGGCGGTGAACCGATGAGCTTTGAGACGGTGTGCGTTCCAT 153
 628 AAGAATACGCCGGGCGGTG 645
 152 GAACCTCACTCATGTGAG 135

Search completed: February 27, 2006, 11:09:29
 Job time : 6438.26 secs

or more than two genes of the *Bacillus subtilis* poly- γ -glutamate synthetase complex (psgBCA) to facilitate microbial surface expression of the target protein. The psgBCA gene complex comprises the pgsB, pgsB and pgsA genes and is normally expressed in the outer membrane of *Bacillus subtilis*. The vector can be transformed into either Gram-positive or Gram-negative bacteria (e.g., *Escherichia coli*), and can be used for the surface expression of various proteins of interest such as enzymes, antigens, antibodies, attachment proteins or adsorption proteins. Proteins recombinantly produced using the vector of the invention can be used as, for example, vaccines or enzymes. The present sequence represents the *Bacillus subtilis* psgBCA complex gene pgsA, which is specifically claimed for use in the vector of the invention.

Sequence 1140 BP; 381 A; 235 C; 265 G; 259 T; 0 U; 0 Other;

Query Match 100.0%; Score 1140; DB 10; Length 1140;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 ATGAAAAAGAACTGAGCTTTCATGAAAAAGCTTAAAGCTGACAAAAAGCAAAAAAAG 60
DB 1 ATGAAAAAGAACTGAGCTTTCATGAAAAAGCTTAAAGCTGACAAAAAGCAAAAAAAG 60
OY 61 AAAACCAATTAAGCAGATTTTATTCATTCGATTCGATTCGATTCGATTCGATTCG 120
DB 61 AAAACCAATTAAGCAGATTTTATTCATTCGATTCGATTCGATTCGATTCGATTCG 120
OY 121 ATGTGGCGGGAAAAAGCGGAAAGCGGAAAGCTTAAAGCTTAAAGCTTAAAGCTT 180
DB 121 ATGTGGCGGGAAAAAGCGGAAAGCGGAAAGCTTAAAGCTTAAAGCTTAAAGCTT 180
OY 181 GCCTCATTTGAGGAGATTTATGATGAGGAGCTTAAAGCTTAAAGCTTAAAGCTT 240
DB 181 GCCTCATTTGAGGAGATTTATGATGAGGAGCTTAAAGCTTAAAGCTTAAAGCTT 240
OY 241 GGGGCGAGCAGATTTTCAATATGTTGAACGATCTTAAAGCTTAAAGCTTAAAGCT 300
DB 241 GGGGCGAGCAGATTTTCAATATGTTGAACGATCTTAAAGCTTAAAGCTTAAAGCT 300
OY 301 GGAACCTTTGAAACCCCGGTAACCTATCAAAAGATTAAACAGCAGATTAAGAGATT 360
DB 301 GGAACCTTTGAAACCCCGGTAACCTATCAAAAGATTAAACAGCAGATTAAGAGATT 360
OY 361 CATCTGACAGCAATTAAGCAATGTAAGTCTTAAAGCTTAAAGCTTAAAGCTTCTC 420
DB 361 CATCTGACAGCAATTAAGCAATGTAAGTCTTAAAGCTTAAAGCTTAAAGCTTCTC 420
OY 421 AACAGCGCAACCAACGAGCAATGATTAAGGCTTCAAGGCGATTAAGATACGCTTGA 480
DB 421 AACAGCGCAACCAACGAGCAATGATTAAGGCTTCAAGGCGATTAAGATACGCTTGA 480
OY 481 GAATTTGGAGCAAAACCTTGATATCGTTGAGGCGGATTAAGCTTAAAGCTTAAAG 540
DB 481 GAATTTGGAGCAAAACCTTGATATCGTTGAGGCGGATTAAGCTTAAAGCTTAAAG 540
OY 541 AAGAAAAATTTGTTACAGAAAGTCAACGCGGTAAGATTAAGCTTAAAGCTTAAAG 600
DB 541 AAGAAAAATTTGTTACAGAAAGTCAACGCGGTAAGATTAAGCTTAAAGCTTAAAG 600
OY 601 GTGTCCGGGAAAGTTTCCGCGCTAAATAAGTAAGCCGCGGCTGCGCGAGATCTT 660
DB 601 GTGTCCGGGAAAGTTTCCGCGCTAAATAAGTAAGCCGCGGCTGCGCGAGATCTT 660
OY 661 GAAATCTTCATCCCTATGATTTTCAAGAGCAAAACATGCTGACATTTGTTGTGAG 720
DB 661 GAAATCTTCATCCCTATGATTTTCAAGAGCAAAACATGCTGACATTTGTTGTGAG 720
OY 721 TCACACTGGGCGCAAGATGATCAATGATTCGCGGCTGCGCGAGATTTGTTGAG 780
DB 721 TCACACTGGGCGCAAGATGATCAATGATTCGCGGCTGCGCGAGATTTGTTGAG 780
OY 781 GGCATGTCTGATGCGGAGCTGACATCATGCTGCGGCTGCGCGAGATTTGTTGAG 840
DB 781 GGCATGTCTGATGCGGAGCTGACATCATGCTGCGGCTGCGCGAGATTTGTTGAG 840

```

```

DB 781 GCCATGTCTGATGCGGAGCTGACATCATGCTGCGGCTGCGCGAGATTTGTTGAG 840
OY 841 ATTAAGTATTAATGAGAACCGGATTTTACAGGCTCGGCAACTTGTCTTGAACCA 900
DB 841 ATTAAGTATTAATGAGAACCGGATTTTACAGGCTCGGCAACTTGTCTTGAACCA 900
OY 901 GGCTGACGAGAACAGACAGTGCATGCTGCTTCAATCACTGAAAGAAATGAGACA 960
DB 901 GGCTGACGAGAACAGACAGTGCATGCTGCTTCAATCACTGAAAGAAATGAGACA 960
OY 961 GGCGGCTTTGAGTGAACAGTATGATTCATGAAAGCACTGCACTGTGAAAAA 1020
DB 961 GGCGGCTTTGAGTGAACAGTATGATTCATGAAAGCACTGCACTGTGAAAAA 1020
OY 1021 GACAGCCTTAAACAGAAACCATTAATTCGGAACCTGACGAAAGACTTAAATTCGTTGG 1080
DB 1021 GACAGCCTTAAACAGAAACCATTAATTCGGAACCTGACGAAAGACTTAAATTCGTTGG 1080
OY 1081 AAAGTAGAAGCGGAAACCTGACGTTGATTAATGATCATGATGACAAACTTAA 1140
DB 1081 AAAGTAGAAGCGGAAACCTGACGTTGATTAATGATCATGATGACAAACTTAA 1140

```

RESULT 2

AD007643
ID AD007643 standard; DNA; 1140 BP.

AC AD007643;

DT 15-JUL-2004 (first entry)

DE B subtilis poly-X-glutamate synthetase complex coding sequence pgsA.

KM ds; gene; enzyme; vaccine; cytosolic; pgsA; poly-X-glutamate synthetase;

KW human papilloma virus.

OS *Bacillus subtilis*.

PN WO2004035795-A1.

PD 29-APR-2004.

PF 17-OCT-2003; 2003WO-KR002163.

PR 17-OCT-2002; 2002KR-00063378.

PA (BIOL-) BIOLEADERS CORP.

PI (KORE-) KOREA RES INST BIOSCIENCE & BIOTECHNOLOG.

PI Sung M, Foo H, Lee J, Jung C, Hong S, Kim C, Park S, Pyo H;

PI WPI: 2004-348463/32.

PT New vector containing pgs A-C genes encoding poly- γ -glutamate

PT synthetase complex and an antigen protein gene of human papilloma virus,

PT useful in preparing vaccine for treating or preventing mucosal tumor,

PT e.g. cervical cancer.

PS Disclosure; Page 62-63; 69pp; English.

XX The present invention relates to a vector for preparing a vaccine which

XX contains one or more than two genes, i.e. pgs A-C encoding poly-X-

XX glutamate synthetase complex and an antigen protein gene of human

XX papilloma virus. The vector and microbes transformed with it are useful

XX in preparing vaccines for treating or preventing mucosal tumor, e.g.

XX cervical cancer. The present sequence is a *Bacillus subtilis* poly-X-

XX glutamate synthetase complex coding sequence.

Sequence 1140 BP; 381 A; 235 C; 265 G; 259 T; 0 U; 0 Other;

Query Match 100.0%; Score 1140; DB 12; Length 1140;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ATGAAAAAGAACTGAGCTTTGATGAAAAAGCTCTAAAGCTGACAAACAGAAAAAAG 60
DB 1 ATGAAAAAGAACTGAGCTTTGATGAAAAAGCTCTAAAGCTGACAAACAGAAAAAAG 60
QY 61 AAAACCAATAGACAGTATTTATGTCATTCGATCGTTTGTCTTATGTTGCTTTC 120
DB 61 AAAACCAATAGACAGTATTTATGTCATTCGATCGTTTGTCTTATGTTGCTTTC 120
QY 121 ATGTGGCGGGAAGAAAGCGGAAACGCCGAAGTCAAAAGCTATTCGACAGCTACTCTCA 180
DB 121 ATGTGGCGGGAAGAAAGCGGAAACGCCGAAGTCAAAAGCTATTCGACAGCTACTCTCA 180
QY 181 GCCTCATTTGTAGCGCATTTATGATGAGGACGCTAATGTTGAAAAAGTAAACGAGCAAAAA 240
DB 181 GCCTCATTTGTAGCGCATTTATGATGAGGACGCTAATGTTGAAAAAGTAAACGAGCAAAAA 240
QY 241 GGGGGCAGACAGTATTTTCAATATGTTGAACCGATCTTTAGAGCTCGGATTAATAGACA 300
DB 241 GGGGGCAGACAGTATTTTCAATATGTTGAACCGATCTTTAGAGCTCGGATTAATAGACA 300
QY 301 GGAATCTTTGAAAAACCCGGTAACTTATCAAAAGAAATTAATAAACAGCAATTAAGAGATT 360
DB 301 GGAATCTTTGAAAAACCCGGTAACTTATCAAAAGAAATTAATAAACAGCAATTAAGAGATT 360
QY 361 CATCTGCAGACGAAATAGAAATCAATGTAAGTCTTGAAGATTTGATTTTCAAGCTTCTC 420
DB 361 CATCTGCAGACGAAATAGAAATCAATGTAAGTCTTGAAGATTTGATTTTCAAGCTTCTC 420
QY 421 AACAGCGCCAAACACCAACCAATGATTCAGGCGTCAAGGCGATGAAGATACGTTGGA 480
DB 421 AACAGCGCCAAACACCAACCAATGATTCAGGCGTCAAGGCGATGAAGATACGTTGGA 480
QY 481 GAATTTGCAAGCAAAACCTTGATATCGTTGAGCGGAGATACAGCTTAAATGATGCGAAA 540
DB 481 GAATTTGCAAGCAAAACCTTGATATCGTTGAGCGGAGATACAGCTTAAATGATGCGAAA 540
QY 541 AAGAAAAATTTCTGACAGAAAGTCAACGCGGATACGATTCGTTAGCGGATTAAGTATGGA 600
DB 541 AAGAAAAATTTCTGACAGAAAGTCAACGCGGATACGATTCGTTAGCGGATTAAGTATGGA 600
QY 601 GTGTCCGGGAAAGGTTTCCGGCTTAAAAAATATCCCGGCGTGCCTCCCGAGATCCCT 660
DB 601 GTGTCCGGGAAAGGTTTCCGGCTTAAAAAATATCCCGGCGTGCCTCCCGAGATCCCT 660
QY 661 GAATCTCTTCAATCCCTATGATTTGAGAAAGCGAAAAAATCATCTGACATTTGTTGTGCA 720
DB 661 GAATCTCTTCAATCCCTATGATTTGAGAAAGCGAAAAAATCATCTGACATTTGTTGTGCA 720
QY 721 TCACACTGGGGCAGAAAGTATGACAAATGATCCAAACGACCGCAGCCGCTTGCAAGA 780
DB 721 TCACACTGGGGCAGAAAGTATGACAAATGATCCAAACGACCGCAGCCGCTTGCAAGA 780
QY 781 GCCATGTTCTATGCGGAGCTGACATATGTTGCGGCAATATCCGCACTGTTAAGAACCG 840
DB 781 GCCATGTTCTATGCGGAGCTGACATATGTTGCGGCAATATCCGCACTGTTAAGAACCG 840
QY 841 ATTGAAGTATATAACGAGACCGTCAATTTTCAACAGCTCCGCAACTTTGCTTTGACCAA 900
DB 841 ATTGAAGTATATAACGAGACCGTCAATTTTCAACAGCTCCGCAACTTTGCTTTGACCAA 900
QY 901 GGGTGCAGAGAACAGAGACAGTGCAGTGGTTCAGTATCACTGAGAAAAATGAGACA 960
DB 901 GGGTGCAGAGAACAGAGACAGTGCAGTGGTTCAGTATCACTGAGAAAAATGAGACA 960
QY 961 GGGCGCTTTGAAGTGAACCGATGATTCATGAAGCAGACTGCACTGTGAAAAAA 1020
DB 961 GGGCGCTTTGAAGTGAACCGATGATTCATGAAGCAGACTGCACTGTGAAAAAA 1020
QY 1021 GAGACGCTTAAACAGAAAAACATTAATTCGGAACCTGACGAAAGAACTCTAATTTGCTTGG 1080
DB 1021 GAGACGCTTAAACAGAAAAACATTAATTCGGAACCTGACGAAAGAACTCTAATTTGCTTGG 1080

```

```

QY 1081 AAAGTAGAAGACGGAAAACTGACGTTTGATATGATCATAGTACAAACTAAATCTTAA 1140
DB 1081 AAAGTAGAAGACGGAAAACTGACGTTTGATATGATCATAGTACAAACTAAATCTTAA 1140

RESULT 3
ADV65734
ADV65734 standard; DNA; 1140 BP.
ADV65734;
10-FEB-2005 (first entry)
DE B. subtilis poly-gamma-glutamic acid synthase A gene SEQ ID NO:3.
KW ds; poly-gamma-glutamic acid synthase A; pgsA; gene expression;
KW antibiotic; antimicrobial; fungicide; cytostatic.
OS Bacillus subtilis.
PN KR2004034780-A.
XX
XX 29-APR-2004.
XX
XX 17-OCT-2002; 2002KR-00063379.
XX
XX 17-OCT-2002; 2002KR-00063379.
XX
XX (BIOL-) BIOLEADERS CORP.
XX (UYCH-) UNIV CHUNGSUN CO LTD.
XX (KORE-) KOREA RES INST BIOSCIENCE & BIOTECHNOLOG.
PI Boo HR, Ham GS, Hong SP, Jung CM, Kim CJ, Lee DG, Lee JS;
PI Park YG, Sung MH;
XX
XX WPI; 2004-577380/56.
PT Method for surface expression of peptides p5 and anal3 using pgs bca
PT gene.
XX
XX Disclosure; SEQ ID NO 3; 25pp; Korean.
XX
XX The invention relates to a novel method for surface expression of
XX peptides p5 and anal3 using a poly-gamma-glutamic acid synthase (pgs) Bca
XX gene, thereby removing a purification process of peptides p5 and anal3,
XX and using lactic acid bacteria for the surface expression, so that
XX peptide antibiotics can be cheaply and stably mass-produced. An
XX expression vector pHClib:pgsA-p5 comprises one or more genes encoding
XX poly-gamma-glutamic acid synthase selected from pgsB, pgsC and pgsA, and
XX a gene for dipolar peptide antibiotics having antimicrobial, antifungal
XX and anticancer activities, wherein the dipolar peptide antibiotic has
XX homology to the peptide p5 encoded by the nucleotide sequence set forth
XX in ADV65735, or to the peptide anal3 encoded by the nucleotide sequence
XX set forth in ADV65737. The present sequence represents the B. subtilis
XX poly-gamma-glutamic acid synthase A gene used in the invention.
XX
XX Sequence 1140 BP; 381 A; 235 C; 265 G; 259 T; 0 U; 0 Other;

Query Match 100.0%; Score 1140; DB 13; Length 1140;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGAAAAAGAACTGAGCTTTGATGAAAAAGCTCTAAAGCTGACAAACAGAAAAAAG 60
DB 1 ATGAAAAAGAACTGAGCTTTGATGAAAAAGCTCTAAAGCTGACAAACAGAAAAAAG 60
QY 61 AAAACCAATAGACAGTATTTATGTCATTCGATCGTTTGTCTTATGTTGCTTTC 120
DB 61 AAAACCAATAGACAGTATTTATGTCATTCGATCGTTTGTCTTATGTTGCTTTC 120
QY 121 ATGTGGCGGGAAGAAAGCGGAAACGCCGAAGTCAAAAGCTATTCGACAGCTACTCTCA 180
DB 121 ATGTGGCGGGAAGAAAGCGGAAACGCCGAAGTCAAAAGCTATTCGACAGCTACTCTCA 180

```

```
QY 181 GCCATTTGTAGGCGATATTATGATGGAGCGTATGTTGAAAAAGTAAAGAGCAAAAA 240
DB 181 GCGTCATTTGTAGGCGATATTATGATGGAGCGTATGTTGAAAAAGTAAAGAGCAAAAA 240
QY 241 GGGGAGAGAGATATTTTCAATATGTTAAACCGATCTTTAGAGCTCGGATTAATGACCA 300
DB 241 GGGGAGAGAGATATTTTCAATATGTTAAACCGATCTTTAGAGCTCGGATTAATGACCA 300
QY 301 GGAACCTTTGAAAAACCCGGTAACTTATCAAAAAGATTTAAACAAGCAATAAAGAGATT 360
DB 301 GGAACCTTTGAAAAACCCGGTAACTTATCAAAAAGATTTAAACAAGCAATAAAGAGATT 360
QY 361 CATCGCAGAGAAATTAAGGAATCAGTGAAGCTTGAAGAGATATGATTAATTCACGGTTCTC 420
DB 361 CATCGCAGAGAAATTAAGGAATCAGTGAAGCTTGAAGAGATATGATTAATTCACGGTTCTC 420
QY 421 AACAGCGCAACCAACCAAGCAATGATTAACGGCTTCAAGGGCATGAAAGATACGCTTGA 480
DB 421 AACAGCGCAACCAACCAAGCAATGATTAACGGCTTCAAGGGCATGAAAGATACGCTTGA 480
QY 481 GAATTTGGGAAAGCAAAACCTTGATATCGTTGAGCGGGATACAGCTTAAAGTATCGGAAA 540
DB 481 GAATTTGGGAAAGCAAAACCTTGATATCGTTGAGCGGGATACAGCTTAAAGTATCGGAAA 540
QY 541 AAGAAATTTTGTATCAGAAAGTCAACGGGGTAAAGATTTGCAAGCGTTGGCTTTACCGAT 600
DB 541 AAGAAATTTTGTATCAGAAAGTCAACGGGGTAAAGATTTGCAAGCGTTGGCTTTACCGAT 600
QY 601 GTGTCCGGGAAAGGTTTTCGCGCTAAAAAGATTAACCGCGGGCGTCTGCGCGAGATCT 660
DB 601 GTGTCCGGGAAAGGTTTTCGCGCTAAAAAGATTAACCGCGGGCGTCTGCGCGAGATCT 660
QY 661 GAAATCTTCATCCCTATGATTTCAAGAGCGAAAAACATGCTGACATTTGTTGTGACG 720
DB 661 GAAATCTTCATCCCTATGATTTCAAGAGCGAAAAACATGCTGACATTTGTTGTGACG 720
QY 721 TCACACTGGGGGCAAGAGATGACAAATATCCAAACGACCGCGACGGCAGCTTTCAGAA 780
DB 721 TCACACTGGGGGCAAGAGATGACAAATATCCAAACGACCGCGACGGCAGCTTTCAGAA 780
QY 781 GCCATGTGTGATGCGGAGCTGACATCATCGCGGCATCATCCGACGCTTTAGAACCG 840
DB 781 GCCATGTGTGATGCGGAGCTGACATCATCGCGGCATCATCCGACGCTTTAGAACCG 840
QY 841 ATTGAAGTATATACGGAACCGTCAATTTTCTACAGCCTTCGCAACTTTGTCTTGAACAA 900
DB 841 ATTGAAGTATATACGGAACCGTCAATTTTCTACAGCCTTCGCAACTTTGTCTTGAACAA 900
QY 901 GGCTGGAGAGAAACAGAGACAGTGCACGTGTTCAAGTATCACTGAAGAAAAATGGAACA 960
DB 901 GGCTGGAGAGAAACAGAGACAGTGCACGTGTTCAAGTATCACTGAAGAAAAATGGAACA 960
QY 961 GGGCGCTTTGAAGTACACCGATCATATCCATGAAGGACACCTGCACTGTGAAAAAA 1020
DB 961 GGGCGCTTTGAAGTACACCGATCATATCCATGAAGGACACCTGCACTGTGAAAAAA 1020
QY 1021 GACAGCCTTTAAACAGAAAAACATTAATTCGCGAATGACGAAAGACTTAAATTCCTTGG 1080
DB 1021 GACAGCCTTTAAACAGAAAAACATTAATTCGCGAATGACGAAAGACTTAAATTCCTTGG 1080
QY 1081 AAAGTGAAGAGGAAAGCTGACGTTGATTAATGATCAATGAGCAAAACTTAAATCTTAA 1140
DB 1081 AAAGTGAAGAGGAAAGCTGACGTTGATTAATGATCAATGAGCAAAACTTAAATCTTAA 1140
```

RESULT 4
ADA44796 standard; DNA; 6536 BP.
AC ADA44796;
XX
XX
DT 20-NOV-2003 (first entry)

```
XX Microbial cell surface expression vector pGNBCA, SEQ ID NO:6.  
DE  
XX  
KW Cell surface; expression vector; microbial;  
KW poly-gamma-glutamate synthetase; Bacillus subtilis pgsBCA complex;  
KW surface expression; Gram-positive bacterium; Gram-negative bacterium;  
KW enzyme; antigen; antibody; attachment protein; adsorption protein;  
KW vaccine; pGNBCA; cyclic; circular; ds.  
XX  
OS Synthetic.  
OS Bacillus subtilis; variety chungkookjang.  
PN  
XX MO2003014360-A1.  
PD  
XX 20-FEB-2003.  
PF  
XX 09-AUG-2002; 2002WO-KR001522.  
PR  
XX 10-AUG-2001; 2001KR-00048373.  
PR  
XX (BIOL-) BIOTRADERS CORP.  
PA (MDMD-) MD LAB CO LTD.  
PI  
XX Sung M, Hong S, Lee J, Jung C, Kim C, Soda K, Ashiuchi M;  
XX WPI; 2003-256589/25.  
DR  
XX  
PT New expression vector containing gene(s) that encode a poly-gamma-  
PT glutamate synthetase complex, useful for producing proteins (e.g.  
PT vaccines or enzymes) on the microbial surface of Gram-positive and/or  
PT Gram-negative bacteria.  
XX  
XX  
PS Example 1; Page 113-120; 122pp; English.  
XX  
CC The invention relates to a vector for expression of a target protein on a  
CC microbial cell surface. The vector of the invention comprises either one  
CC or more than two genes of the Bacillus subtilis poly-gamma-glutamate  
CC synthetase complex (pgsBCA) to facilitate microbial surface expression of  
CC the target protein. The pgsBCA gene complex comprises the pgsB, pgsB and  
CC pgsA genes and is normally expressed in the outer membrane of Bacillus  
CC subtilis. The vector can be transformed into either Gram-positive or Gram  
CC -negative bacteria (e.g., Escherichia coli), and can be used for the  
CC surface expression of various proteins of interest such as enzymes,  
CC antigens, antibodies, attachment proteins or adsorption proteins.  
CC Proteins recombinantly produced using the vector of the invention can be  
CC used as, for example, vaccines or enzymes. The present sequence  
CC represents a microbial cell surface expression vector of the invention,  
CC pGNBCA.  
XX  
SQ Sequence 6536 BP; 1855 A; 1464 C; 1567 G; 1649 T; 0 U; 1 Other;  
Query Match 100.0%; Score 1140; DB 10; Length 6536;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGAAAAAAGAACTGAGCTTTCATGAAAAAGCTGTAAGCTGACAAAAACAGCAAAAAAG 60  
DB 1891 ATGAAAAAAGAACTGAGCTTTCATGAAAAAGCTGTAAGCTGACAAAAACAGCAAAAAAG 1950  
QY 61 AAAACCAATGAAGCAATATTTATGCAATTCGATCGATGCTTTTGTCTTATGTTGCTTTTC 120  
DB 1951 AAAACCAATGAAGCAATATTTATGCAATTCGATCGATGCTTTTGTCTTATGTTGCTTTTC 2010  
QY 121 ATGTGGGGGGGAAAAAGCGGAAACGCCCAAGTCAAAACGTTTGTGACGACGTACTCTCA 180  
DB 2011 ATGTGGGGGGGAAAAAGCGGAAACGCCCAAGTCAAAACGTTTGTGACGACGTACTCTCA 2070  
QY 181 GCCTCATTTGTAGGCGATATTATGATGGAGCGTATGTTGAAAAAGTAAAGAGCAAAAA 240  
DB 2071 GCGTCATTTGTAGGCGATATTATGATGGAGCGTATGTTGAAAAAGTAAAGAGCAAAAA 2120  
QY 241 GGGGAGAGAGATATTTTCAATATGTTAAACCGATCTTTAGAGCTCGGATTAATGACCA 300
```

Db 2131 GGGGAGACATATTTTCAATATGTTGACCGATCTTAGAGCTCGGATATATGACA 2130
 QY 301 GGAACCTTTGAAAAACCGGTAACTATCAAAAGATTATTAACAAGCAGATTAAGAGATT 360
 Db 2191 GGAACCTTTGAAAAACCGGTAACTATCAAAAGATTATTAACAAGCAGATTAAGAGATT 2250
 QY 361 CATCTGCAGAGATTAAGAGATTCAGTGAAGCTTTGAAAGGATTAATTTACGGTTCTC 420
 Db 2251 CATCTGCAGAGATTAAGAGATTCAGTGAAGCTTTGAAAGGATTAATTTACGGTTCTC 2310
 QY 421 AACGCGCCCAACACCAAGCATGATTAACGGCTTTAGGGCATGAAGATACGTTTGA 480
 Db 2311 AACGCGCCCAACACCAAGCATGATTAACGGCTTTAGGGCATGAAGATACGTTTGA 2370
 QY 481 GAATTTGGGAAACAAACCTTGATATCGTTGAGCGGAGATACAGTTAAGTATCGGAA 540
 Db 2371 GAATTTGGGAAACAAACCTTGATATCGTTGAGCGGAGATACAGTTAAGTATCGGAA 2430
 QY 541 AAGAAAATTTGATACAGAAAGTCAACGGGTTAAGATTTGCAACGCTTGCTTACGAT 600
 Db 2431 AAGAAAATTTGATACAGAAAGTCAACGGGTTAAGATTTGCAACGCTTGCTTACGAT 2490
 QY 601 GTGTCGGGAAAGGTTTGGCGGCTAAAAAGAAATACGCGGCGGTCTGCGCCAGATCT 660
 Db 2491 GTGTCGGGAAAGGTTTGGCGGCTAAAAAGAAATACGCGGCGGTCTGCGCCAGATCT 2550
 QY 661 GAAATCTTCACTCCCTATGATTTAGAAAGGAAAAACATGCTGACATTTGTTTGCGAG 720
 Db 2551 GAAATCTTCACTCCCTATGATTTAGAAAGGAAAAACATGCTGACATTTGTTTGCGAG 2610
 QY 721 TCACACTGGGCGCAAGATGATGACATGATCCAAACGACCGGCGGCGGCGGCGGCGG 780
 Db 2611 TCACACTGGGCGCAAGATGATGACATGATCCAAACGACCGGCGGCGGCGGCGGCGG 2670
 QY 781 GGCATGTCGATGCGGAGGCTGACATCATCGTGGCGCATCATCCGACGCTTTAGAACG 840
 Db 2671 GGCATGTCGATGCGGAGGCTGACATCATCGTGGCGCATCATCCGACGCTTTAGAACG 2730
 QY 841 ATTGAAGTATATACGGAACCGTCAATTTTCTACAGCTTGGGCACTTTGTCTTGAACAA 900
 Db 2731 ATTGAAGTATATACGGAACCGTCAATTTTCTACAGCTTGGGCACTTTGTCTTGAACAA 2790
 QY 901 GGCTGACGAGAAACAAGACAGTGCATGCTGTTCACTATCACTGAAGAAAATGGAACA 960
 Db 2791 GGCTGACGAGAAACAAGACAGTGCATGCTGTTCACTATCACTGAAGAAAATGGAACA 2850
 QY 961 GGGCGCTTTGAAGTACACCGATTCATCCATGAAAGGACACCTGCACTGTGAAGAAA 1020
 Db 2851 GGGCGCTTTGAAGTACACCGATTCATCCATGAAAGGACACCTGCACTGTGAAGAAA 2910
 QY 1021 GACAGCCTTAAACAGAAAACATTTATTCGGAAGTACGAAAGAGCTCTAATTTCCGTTGG 1080
 Db 2911 GACAGCCTTAAACAGAAAACATTTATTCGGAAGTACGAAAGAGCTCTAATTTCCGTTGG 2970
 QY 1081 AAGTAGAAGACGAAAGCTGATTTGATTTGATGATGATGACAACTAAATCTAAA 1140
 Db 2971 AAGTAGAAGACGAAAGCTGATTTGATTTGATGATGATGACAACTAAATCTAAA 3030

RESULT 5
 AAF82257
 ID AAF82257 standard; DNA; 1143 BP.

AAF82257;

21-JUN-2001 (first entry)

Bacillus subtilis IFO 3336 DNA encoding a PGA synthesizing enzyme.

Bacterial; poly-gamma-L-glutamic acid; L-PGA;

KM poly-gamma-D,L-glutamic acid; PGA; glutamate racemase; ds.

XX
XX
OS Bacillus subtilis.

XX Key Location/Qualifiers
 FH CDS 1..1143
 FT /*tag= a
 JP2001017182-A.
 XX 23-JAN-2001.
 XX 09-JUL-1999; 99JP-00196335.
 XX 09-JUL-1999; 99JP-00196335.
 XX 09-JUL-1999; 99JP-00196335.
 XX (NAGS) NAGASE SANGYO KK.
 XX WPI: 2001-285408/30.
 XX P-PSDB; MAB74026.
 PT New nucleic acid encoding a glutamate racemase enzyme useful for the
 PT preparation of poly-gamma-glutamic acid.
 XX Claim 4; Page 13-15; 17pp; Japanese.
 XX The present sequence encodes an enzyme which is useful in the production
 CC of poly-gamma-L-glutamic acid (L-PGA) or poly-gamma-D,L-glutamic acid
 CC (PGA). A plasmid comprising the present sequence may be used to transform
 CC Escherichia coli. The transformants express the enzyme and PGA is
 CC produced in the culture
 SQ Sequence 1143 BP; 385 A; 233 C; 264 G; 261 T; 0 U; 0 Other;

Query Match 99.3%; Score 1132; DB 4; Length 1143;
 Best Local Similarity 99.6%; Pred. No. 0; Mismatches 5; Indels 0; Gaps 0;

Matches 1135; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ATGAAAAAAGAACTGAGCTTTTCATGAAAGCTGTAAGCTGACAAAACAGCAAAAAAAG 60
 Db 1 ATGAAAAAAGAACTGAGCTTTTCATGAAAGCTGTAAGCTGACAAAACAGCAAAAAAAG 60
 QY 61 AAAACCAATAGACAGATTTATTTATTCGATTCGATCGTTTGTCTTATGTTCCGCTTC 120
 Db 61 AAAACCAATAGACAGATTTATTTATTCGATTCGATTCGATCGTTTGTCTTATGTTCCGCTTC 120
 QY 121 ATGTGGGCGGAAAAAGCGGAAAAAGCGGAAAGTCAAAAGTATTCGACGACGCTACTTCA 180
 Db 121 ATGTGGGCGGAAAAAGCGGAAAAAGCGGAAAGTCAAAAGTATTCGACGACGCTACTTCA 180
 QY 181 GCTCATTTGTAGGAGATATTTATGATGGGACGCTATGTTGAAGAAAAGTAAAGGCAAAA 240
 Db 181 GCTCATTTGTAGGAGATATTTATGATGGGACGCTATGTTGAAGAAAAGTAAAGGCAAAA 240
 QY 241 GGGGAGACAGTATTTTCAATATGTTGAACGATCTTTAGAGCTCGATTTATGATGACA 300
 Db 241 GGGGAGACAGTATTTTCAATATGTTGAACGATCTTTAGAGCTCGATTTATGATGACA 300
 QY 301 GGAACCTTTGAAAACCCGGTAACTATCAAAAGATTTATTAACAAGCAGATAAGAGATT 360
 Db 301 GGAACCTTTGAAAACCCGGTAACTATCAAAAGATTTATTAACAAGCAGATAAGAGATT 360
 QY 361 CATCTGCAGAGATTAAGAGATTCAGTGAAGCTTTGAAAGGATATGATTTCAAGGTTCTC 420
 Db 361 CATCTGCAGAGATTAAGAGATTCAGTGAAGCTTTGAAAGGATATGATTTCAAGGTTCTC 420
 QY 421 AACAGCGCCCAACACCAAGCATGATTAACGGCTTTAGGGCATGAAGATACGTTTGA 480
 Db 421 AACAGCGCCCAACACCAAGCATGATTAACGGCTTTAGGGCATGAAGATACGTTTGA 480
 QY 481 GAATTTGGGAAACAAACCTTGATATCGTTGAGCGGAGATACAGTTAAGTATGAGGAAA 540
 Db 481 GAATTTGGGAAACAAACCTTGATATCGTTGAGCGGAGATACAGTTAAGTATGAGGAAA 540
 QY 541 AAGAAAATTTGATACAGAAAGTCAACGGGTTAAGATTTGCAACGCTTGCTTACGAT 600
 Db 541 AAGAAAATTTGATACAGAAAGTCAACGGGTTAAGATTTGCAACGCTTGCTTACGAT 600

```
Db 541 AAGAAATTTCTGATACAAAAGTCACGGGGTAACGATTTGGACGCTTGCTTACCGAT 600
Qy GTGTCGGGGAAGGTTTTCGGGCTTAAAGAAATCGCCGGCGTGTCTCCGCGAGATCCT 660
Db 601 GTGTCGGGGAAGGTTTTCGGGCTTAAAGAAATACCGGGCGTGTCTCCGCGAGATCCT 660
Qy 661 GAAATCTTTCATCCCTATGATTTTCAGAACCGGAAAAAACAATGTGACATTTGTTGTGTGAG 720
Db 661 GAAATTTTCATCCCTATGATTTTCAGAACCGGAAAAAACAATGTGACATTTGTTGTGTGAG 720
Qy 721 TCACACTGGGGCCCAAGAGTATGACATGATCCAAACGACCGCCAGCCGACGTTGCAAGA 780
Db 721 TCACACTGGGGCCCAAGAGTATGACATGATCCAAACGACCGCCAGCCGACGTTGCAAGA 780
Qy 781 GCCATGTCGATGGGGGAGTGAATCATCGTGGGCGCATCATCCGACGTCCTTGAACCG 840
Db 781 GCCATGTCGATGGGGGAGTGAATCATCGTGGGCGCATCATCCGACGTCCTTGAACCG 840
Qy 841 ATTGAAGTATATTAACGGAACCGTCATTTTCTACAGCCCTCGGCACTTTGTCTTGACCAA 900
Db 841 ATTGAAGTATATTAACGGAACCGTCATTTTCTACAGCCCTCGGCACTTTGTCTTGACCAA 900
Qy 901 GCGTGGACGAGAACGAGACAGTGCACCTGTTCAATCACTGGAAGAAAAATGGAACA 960
Db 901 GCGTGGACGAGAACGAGACAGTGCACCTGTTCAATCACTGGAAGAAAAATGGAACA 960
Qy 961 GCGCGCTTTGAAGTACACCGATCGATTCATGTAAGGACACCTGCACTGTGAAAAA 1020
Db 961 GCGCGCTTTGAAGTACACCGATCGATTCATGTAAGGACACCTGCACTGTGAAAAA 1020
Qy 1021 GACAGCCTTTAAACAGAAAACCTTATTCGCGAATCGAGAAAAGACTTAATTTGCTTTGG 1080
Db 1021 GACAGCCTTTAAACAGAAAACCTTATTCGCGAATCGAGAAAAGACTTAATTTGCTTTGG 1080
Qy 1081 AAAGTAGAAGACGGAACCTGACGTTTGAATGATGATGACAAACTAAATCTAAA 1140
Db 1081 AAAGTAGAAGACGGAACCTGACGTTTGAATGATGATGATGACAAACTAAATCTAAA 1140

RESULT 6
AAF82254
ID AAF82254 standard; DNA; 3045 BP.
XX
AC AAF82254;
XX
DT 21-JUN-2001 (first entry)
XX
DE Bacillus subtilis IPO 3336 DNA encoding glutamate racemase enzyme.
XX
KW Bacterial; poly-gamma-L-glutamic acid; L-PGA;
XX
KW poly-gamma-D,L-glutamic acid; PGA; glutamate racemase; ds.
XX
OS Bacillus subtilis.
XX
FH Key Location/Qualifiers
FT CDS 19..1200
FT CDS /*tag= a
FT CDS 1215..1264
FT CDS /*tag= b
FT CDS 1683..2825
FT CDS /*tag= c
XX
PN JP2001017182-A.
XX
PD 23-JAN-2001.
XX
PF 09-JUL-1999; 99JP-00196335.
XX
PR 09-JUL-1999; 99JP-00196335.
XX
PA (NAGS ) NAGASE SANGYO KK.
XX
DR WPI; 2001-285408/30.
```

```
DR P-PSDB; AAB74024, AAB74025, AAB74026.
XX
PT New nucleic acid encoding a glutamate racemase enzyme useful for the
PT preparation of poly-gamma-glutamic acid.
XX
PS Claim 1; Page 7-11; 17pp; Japanese.
XX
CC The present sequence encodes a glutamate racemase enzyme which is useful
CC in the production of poly-gamma-L-glutamic acid (L-PGA) or poly-gamma-D,L
CC -glutamic acid (PGA). A plasmid comprising the present sequence may be
CC used to transform Escherichia coli. The transformants express the enzyme
CC and PGA is produced in the culture
XX
SQ Sequence 3045 BP; 1000 A; 593 C; 696 G; 756 T; 0 U; 0 Other;
Query Match 99.3%; Score 1132; DB 4; Length 3045;
Beet Local Similarity 99.6%; Pred. No. 0;
Matches 1135; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 1 ATGAAAAAAGAACTGACCTTTCATGAAAAGCTGTAAGCTGACAAAACAGCAAAAAAG 60
Db 1683 ATGAAAAAAGAACTGACCTTTCATGAAAAGCTGTAAGCTGACAAAACAGCAAAAAAG 1742
Qy 61 AAAACCAATTAAGCAGTATTTATTCGCAATTCGATCGTTTGTCTTATGTTGCTTTTC 120
Db 1743 AAAACCAATTAAGCAGTATTTATTCGCAATTCGATCGTTTGTCTTATGTTGCTTTTC 1802
Qy 121 ATGTGGCGGGGAAAAAGCGGAAACGCGAAGTCAAAACGATTCCTGACGAGTACTCTCA 180
Db 1803 ATGTGGCGGGGAAAAAGCGGAAACGCGAAGTCAAAACGATTCCTGACGAGTACTCTCA 1862
Qy 181 GCCTCATTTGTAAGCGATATTTATGATGAGGACGCTATGTTGAAAAAGTACGAGCAAAAA 240
Db 1863 GCCTCATTTGTAAGCGATATTTATGATGAGGACGCTATGTTGAAAAAGTACGAGCAAAAA 1922
Qy 241 GGGGCAAGCAGTATTTTCAATATGTTGAAACCGATCTTGAAGCTTCGATATGTAACA 300
Db 1923 GGGGCAAGCAGTATTTTCAATATGTTGAAACCGATCTTGAAGCTTCGATATGTAACA 1982
Qy 301 GGAACCTTTGAAAAACCGGTTAACCTATCAAAAGAAATTAATAACAAGAGATTAAGAGATT 360
Db 1983 GGAACCTTTGAAAAACCGGTTAACCTATCAAAAGAAATTAATAACAAGAGATTAAGAGATT 2042
Qy 361 CATCTGCAGACGATTAAGAAATCAGTAAAGTCTTGAAGGATATGATTTTACCGTTCTC 420
Db 2043 CATCTGCAGACGATTAAGAAATCAGTAAAGTCTTGAAGGATATGATTTTACCGTTCTC 2102
Qy 421 AACAGCGCAACCAACGCAATGATTAACGCGTTCAAGGCGATGAAGATACGCTTGA 480
Db 2103 AACAGCGCAACCAACGCAATGATTAACGCGTTCAAGGCGATGAAGATACGCTTGA 2162
Qy 481 GAATTTGCGAAGCAAAACCTTGATATGCTTGACCGGATTAACGCTTAAGTATGAGCAAA 540
Db 2163 GAATTTGCGAAGCAAAACCTTGATATGCTTGACCGGATTAACGCTTAAGTATGAGCAAA 2222
Qy 541 AAGAAAAATTTCTGATACAAAAGTCAACGGGGTAAAGATTGCAAGCTTGGCTTACCGAT 600
Db 2223 AAGAAAAATTTCTGATACAAAAGTCAACGGGGTAAAGATTGCAAGCTTGGCTTACCGAT 2282
Qy 601 GTGTCCGGGAAAGGTTTCGCGGCTTAAAGAAATACGCGCGGCGTGTCTCCGCGAGATCCT 660
Db 2283 GTGTCCGGGAAAGGTTTCGCGGCTTAAAGAAATACGCGCGGCGTGTCTCCGCGAGATCCT 2342
Qy 661 GAAATCTTTCATCCCTATGATTTTCGAAGCGGAAAAAACAATGTGACATTTGTTGTGTGAG 720
Db 2343 GAAATCTTTCATCCCTATGATTTTCGAAGCGGAAAAAACAATGTGACATTTGTTGTGTGAG 2402
Qy 721 TCACACTGGGGCCCAAGAGTATGACATGATCCAAACGACCGCCAGCCGACGTTGCAAGA 780
Db 2403 TCACACTGGGGCCCAAGAGTATGACATGATCCAAACGACCGCCAGCCGACGTTGCAAGA 2462
Qy 781 GCCATGTCGATGGGGGAGTGAATCATCGTGGGCGCATCATCCGACGTCCTTGAACCG 840
```

Db 2463 GCCATGTCGATGCGGAGAGCTGATCATCTGCGCCATCATCCGACGCTTTAGAACG 2522
 QY 841 ATTGAAGTATATAACGAAACCGTCATTTTCTACAGCCTCGCACTTTGCTTTGACCA 900
 Db 2523 ATTGAAGTATATAACGAAACCGTCATTTTCTACAGCCTCGCACTTTGCTTTGACCA 2582
 QY 901 GCGTGCAGAGAACAGACAGACAGTCTGCTTCAGTCTGCTTCAGTCTGCTTCAGTCTG 960
 Db 2583 GCGTGCAGAGAACAGACAGACAGTCTGCTTCAGTCTGCTTCAGTCTGCTTCAGTCTG 2642
 QY 961 GCGGCGCTTTGAGTACACCGATGATATCCATGAGGACACCTGACCTGTGAAAAA 1020
 Db 2643 GCGGCGCTTTGAGTACACCGATGATATCCATGAGGACACCTGACCTGTGAAAAA 2702
 QY 1021 GACAGCCTTTAAACAGAAAAACCTATTATTCGCGAAGTACGAAAGCTTAAATTTGCTTGG 1080
 Db 2703 GACAGCCTTTAAACAGAAAAACCTATTATTCGCGAAGTACGAAAGCTTAAATTTGCTTGG 2762
 QY 1081 AAGTGAAGACGAGAAACCTGACGTTTGTATTTGATGATGATGATGATGATGATGATGAT 1140
 Db 2763 AAGTGAAGACGAGAAACCTGACGTTTGTATTTGATGATGATGATGATGATGATGATGAT 2822

RESULT 7

ABK74454
 ID ABK74454 standard; DNA; 696 BP.

ABK74454;

13-AUG-2002 (first entry)

Bacillus licheniformis genomic sequence tag (GST) #1745.

Differential gene expression; genomic sequenced tag; GST; altered culture condition; environmental stress; physiological provocation; ds.

Bacillus licheniformis.

WO200229113-A2.

11-APR-2002.

05-OCT-2001; 2001WO-US031437.

06-OCT-2000; 2000US-00680598.

27-MAR-2001; 2001US-0279526P.

(NOVO) NOVOZYMES BIOTECH INC.

(NOVO) NOVOZYMES AS.

Berka R, Clausen IG;

WPI; 2002-41684/44.

Monitoring differential expression of several genes in first Bacillus cell relative to expression of same genes in one or more second Bacillus cells, by using substrate containing Bacillus genomic sequenced tag array.

Claim 4; SEQ ID NO 1745; 200pp; English.

The invention describes a method of monitoring differential expression of genes in a first Bacillus cell relative to expression of the genes in other Bacillus cells, comprising hybridising labeled nucleic acid probes isolated from Bacillus cells to a substrate containing array of Bacillus genomic sequenced tags (GST), examining the array, and determining relative gene expression by an observed hybridisation reporter signal of a spot in the array. The method is useful for measuring the expression of genes in a first Bacillus cell relative to expression of the same genes in one or more second Bacillus cells. The method is useful for monitoring global expression of several genes from a Bacillus cell, discovering new genes, identifying possible functions of unknown open reading frames and

CC monitoring gene copy number variation and stability. Monitoring changes in expression of genes may be used to provide a representation of the way in which Bacillus cells adapt to changes in culture conditions, environmental stress or other physiological provocation. Extensive follow-up characterisation is unnecessary, when one spot on an array equals one gene or one open reading frame, since sequence information is available. CC This sequence represents a genomic sequence tag (GST) used in the method of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at http://wipo.int/pub/published_pct_sequences

Sequence 696 BP; 226 A; 151 C; 173 G; 146 T; 0 U; 0 Other;

Query Match 29.2%; Score 332.4; DB 6; Length 696;

Best Local Similarity 68.4%; Pred. No. 7.1e-85;

Matches 475; Conservative 0; Mismatches 216; Indels 3; Gaps 1;

QY 444 GGATTAACGCGCTTCAGGCGATGAAAGATACCTTGAGAAATTTGGAGAAACCTTTGA 503
 Db 1 GGACTACGGGGAAGACGGTTTGAAGATACCTCAATTAATTTCAATGAGAAATCTGA 60
 QY 504 TATCGTGCAGCGGATACAGCTTAAGTATGCGAAAGAAATTTGTAACGGAAGT 563
 Db 61 GCTTGTCCGAGCAGGAAATTAATTTGAAGACGCGAAACAGACGATCTATCGAATGT 120
 QY 564 CAACGGGTTACGATTTGCAAGCGTTGCTTACCGATGTCGCGGAAAGTTTGCAGC 623
 Db 121 GAACGGGTTAAATTTGCAAGCGTCTGCTTTACAGACGTTACACAAAGAACTTACAGC 180
 QY 624 TAAAGAAATACGCGCGCGCTGCTGCCGAGATCTGAAATCTTATCTTATGATTTG 683
 Db 181 CAAAGAAACAGAGCGGAGAGTCTGCCCTCAGTCGAAATCTTATTTCAATGATTC 240
 QY 684 AGAACGAAAAAATGCTGATGATTTGTTGTCAGTCACTGCGGCGCAAGATTA 743
 Db 241 GGAGCATGCAAAAAAGCGGATCTTGTCTTGTCCATGTCAGTGGGACAAAGATTA 300
 QY 744 CAATGATCCAAACGACCGCGCGGCTTGCAGAGCAATGTCAGTGGGAGCTGA 803
 Db 301 CAATGATCCGAAACGACGACGAGATCTGCGCAAGGAGATGCAATGCGGAGCAGA 360
 QY 804 CATCATCGTGGCGCATCATCCGACGCTTTAGAACCGATTTGAATATTAACGGAACCT 863
 Db 361 TGTCAATCATGGCGCTCATCCCATGTTCTGCAACGATCGAAGTATTAACGTAAGT 420
 QY 864 CATTTTCTACAGCCTCGCACTTTGCTTTGACCAAGGCTGACGAGACAGACAG 923
 Db 421 GATTTTCTACAGCCTCGCACTTTGATGATGATGATGATGATGATGATGATGATGAT 480
 QY 924 TGCACGCTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 983
 Db 481 CGCGCTTGTACAAATCCATTTAATGAATGACGCAAGGCGCTTTGAGTAAAGCTCT 540
 QY 984 CGATATCATGAGAGACACCTGACCT--GTGAAAAAAGACAGCTTTAAACAGAAAC 1040
 Db 541 CAACATTCGCGAAGACGCGCGACGCTTTAGCAAGGCGACTTTTAAAGAAAGC 600
 QY 1041 CATTTATTCGCGAAGTACGAAAGACTTAAATTTGCTTGAAGTGAAGAGGAAACT 1100
 Db 601 GATCTTCGCTCAATGACAAAGAAACAACTCGACTGGAAGAGAGAAAGGAAATTT 660
 QY 1101 GACGTTGATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1134
 Db 661 AACGTTGAAGTCATGCGAGCAAGCTGAA 694

RESULT 8

ABK74486

ABK74486;

13-AUG-2002 (first entry)

XX	Bacillus licheniformis genomic sequence tag (GST) #1777.
DE	
XX	
KW	Differential gene expression; genomic sequenced tag; GST;
KW	altered culture condition; environmental stress;
KW	physiological provocation; ds.
XX	
OS	Bacillus licheniformis.
XX	
PN	WO200229113-A2.
XX	
PD	11-APR-2002.
XX	
PF	05-OCT-2001; 2001WO-US031437.
XX	
PR	06-OCT-2000; 2000US-00680598.
PR	27-MAR-2001; 2001US-0279526P.
XX	
XX	
PA	(NOVO) NOVOZYMES BIOTECH INC.
PA	(NOVO) NOVOZYMES AS.
XX	
PI	Berka R, Clausen IG;
XX	
DR	WPI; 2002-416684/44.
XX	
PT	Monitoring differential expression of several genes in first Bacillus
PT	cell relative to expression of same genes in one or more second Bacillus
PT	cells, by using substrate containing Bacillus genomic sequenced tag
PT	array.
XX	
XX	
XX	Claim 4; SEQ ID NO 1777; 200dp; English.

CC The invention describes a method of monitoring differential expression of
CC genes in a first *Bacillus* cell relative to expression of the genes in
CC other *Bacillus* cells, comprising hybridizing labelled nucleic acid probes
CC isolated from *Bacillus* cells to a substrate containing array of *Bacillus*
CC genomic sequenced tags (GST), examining the array, and determining
CC relative gene expression by an observed hybridisation reporter signal of
CC a spot in the array. The method is useful for measuring the expression of
CC genes in a first *Bacillus* cell relative to expression of the same genes
CC in one or more second *Bacillus* cells. The method is useful for monitoring
CC global expression of several genes from a *Bacillus* cell, discovering new
CC genes, identifying possible functions of unknown open reading frames and
CC monitoring gene copy number variation and stability. Monitoring changes
CC in expression of genes may be used to provide a representation of the way
CC in which *Bacillus* cells adapt to changes in culture conditions,
CC environmental stresses or other physiological provocation. Extensive follow
CC up characterisation is unnecessary, when one spot on an array equals one
CC gene or one open reading frame, since sequence information is available.
CC This sequence represents a genomic sequence tag (GST) used in the method
CC of the invention. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from Wipo at http://wipo.int/pub/published_pat_sequences
XX

Sequence 488 BP; 178 A; 89 C; 105 G; 116 T; 0 U; 0 Other;

Query Match	19.8%	Score 225.6	DB 6	Length 488
Best Local Similarity	67.8%	Pred. No. 3.3e-54		
Matches 331, Conservative	0	Mismatches 154	Indels 3	Gaps 1

QY	1	ATGAAAAAAAAAGAACTGAGCTTTTCATGAAAAGCTGCTAAAGCTGACAAACAGCAAAAAAAG	60
Db	1	ATGAAAAAAAACAACTGAACTTTTCAGGAAAAAATGCTGAAAGTTGACGAAGCAGGAGAAAAAG	60
QY	61	AAACCAATAAGCAGATTTATTTATTCGCATTCGATGCTTTTGTCCCTAATGTTGCGCTTC	120
Db	61	AAAAACAACACACGCTCTTTATTCGATTTGCCGATTCCTGTTAATGTTGCTTT	120
QY	121	ATGTGGCGCGGAAAAAGCGAAAAAGC---CGAAGGTCAAAACGATTTCTGACGACGATCTC	177
Db	121	ACTTGGGTGGAAGCCCAAAACTCTTTCGCAAAATGACAAAAAAGAAAGTCCAAAGCTT	180
QY	178	TCAGGCTCATTTGTGAGCGATATTAATGATGGACGCTAATGTTGAAAAAGTAAACGAGCA	237

Db	181	ACAGCTACTTTTGGTGGCATATCATGATGGGAAAGAAACGTAGAAAAAGTGACAAACTTG	240
Qy	238	AAAGGGGCAACAGTATTTTTCATATGTTGAACCGATCTTTAGAGCCTCGGATTATGTA	297
Db	241	CACGGTTCGGAAAGTGCTCTTCAAAAATGTGAAGCCGTACTTTAATGTGCAGATTTTATC	300
Qy	298	GCAGGAACTTTGAAAAACCCGGTAACATCAAAAAGATTTAAACAAGCAGATAAAGAG	357
Db	301	ACAGGAACTTTGAAAAACCCGTGAACCAATGCAGAAAGACTTACAGAGGCCAGAAAGAAC	360
Qy	358	ATTTCATCTGCAGACGATAGGAATCAGTGAAGCTTTGAAGATATGAATTTTACCGTT	417
Db	361	ATCATCTGTCAAAACGATCAAGATCAATCAAGTCAATTTGAAAAAGCTGAACTTCAGCGTA	420
Qy	418	CTCAACAGGCCCAACACCGCAATGGAATTCGCGCTTCAGGGCATGAAAAGATACGCTT	477
Db	421	CTGAATTTTTCCAACCAACCATCGATGACTACGGGGAAGACGGCTTTGAAAGGATACGCTC	480
Qy	478	GGAGAAATT	485
Db	481	AATGAGTT	488

RESULT 9
ADB06115
ID ADB06115 standard; DNA; 1212 BP.

AC ADB06115;

DT 20-NOV-2003 (first entry)

DE **Aloiococcus** otitis antigenic protein encoding DNA SEQ ID NO:55.

Alloiococcus otitidis; antigenic protein; immunogenic; immunisation; KW
KW
gene therapy; Gram-positive bacterium; infection; gene; ds.

Alloiococcus otitis.

PN WO2003048304-A2

PD 12-JUN-2003

PF 25-NOV-2002; 2002WO-US036123.

PR 29-NOV-2001; 2001US-0333777P.

XX	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	51	54	57	60	63	66	69	72	75	78	81	84	87	90	93	96	99	102	105	108	111	114	117	120	123	126	129	132	135	138	141	144	147	150	153	156	159	162	165	168	171	174	177	180	183	186	189	192	195	198	201	204	207	210	213	216	219	222	225	228	231	234	237	240	243	246	249	252	255	258	261	264	267	270	273	276	279	282	285	288	291	294	297	300	303	306	309	312	315	318	321	324	327	330	333	336	339	342	345	348	351	354	357	360	363	366	369	372	375	378	381	384	387	390	393	396	399	402	405	408	411	414	417	420	423	426	429	432	435	438	441	444	447	450	453	456	459	462	465	468	471	474	477	480	483	486	489	492	495	498	501	504	507	510	513	516	519	522	525	528	531	534	537	540	543	546	549	552	555	558	561	564	567	570	573	576	579	582	585	588	591	594	597	600	603	606	609	612	615	618	621	624	627	630	633	636	639	642	645	648	651	654	657	660	663	666	669	672	675	678	681	684	687	690	693	696	699	702	705	708	711	714	717	720	723	726	729	732	735	738	741	744	747	750	753	756	759	762	765	768	771	774	777	780	783	786	789	792	795	798	801	804	807	810	813	816	819	822	825	828	831	834	837	840	843	846	849	852	855	858	861	864	867	870	873	876	879	882	885	888	891	894	897	900	903	906	909	912	915	918	921	924	927	930	933	936	939	942	945	948	951	954	957	960	963	966	969	972	975	978	981	984	987	990	993	996	999
XX	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	51	54	57	60	63	66	69	72	75	78	81	84	87	90	93	96	99	102	105	108	111	114	117	120	123	126	129	132	135	138	141	144	147	150	153	156	159	162	165	168	171	174	177	180	183	186	189	192	195	198	201	204	207	210	213	216	219	222	225	228	231	234	237	240	243	246	249	252	255	258	261	264	267	270	273	276	279	282	285	288	291	294	297	300	303	306	309	312	315	318	321	324	327	330	333	336	339	342	345	348	351	354	357	360	363	366	369	372	375	378</																																																																																																																																																																																																															

XX

XX

DR P-PSDB; ADB06116.

PT New Alliococcus otitidis po

PT effects during drug clinical

PS Claim 7; SEQ ID NO 55; 1019E

CC The present invention describes

CC *Alloicoccus otitidis* is a C

CC expression vector comprising

CC host cell, transfected, trans

CC composition comprising the K

CC expression vector; (6) a phage

100

PS Claim 7, SEQ ID NO 55, 101pp; English.

of the polypeptides of (1), their biological equivalent or fragment; (8) immunising against *Allotiococcus ostitidis* by administering to a host the immunogenic composition; (9) detecting and/or identifying *Allotiococcus ostitidis* in the biological sample; (10) a kit comprising a container containing the novel polynucleotide, its degenerate variant or fragment, or the antibody of (4); and (11) producing a polypeptide by culturing the genetically engineered host cell under conditions suitable to produce the polypeptide from the culture. (1) can be used in gene therapy. The polynucleotides, polypeptides, antibodies and compositions of the present invention can be used for treating and diagnosing diseases, drug screening assays and monitoring of effects during drug clinical trials. The polynucleotides are useful for expressing and detecting *Allotiococcus ostitidis*. The present sequence encodes an *Allotiococcus ostitidis* antigen protein from the present invention.

Sequence 1212 BP; 353 A; 248 C; 259 G; 352 T; 0 U; 0 Other;

Query Match 8.5%; Score 96.4; DB 9; Length 1212;
Best Local Similarity 46.7%; Pred. No. 7.8e-17;
Matches 393; Conservative 0; Mismatches 431; Indels 18; Gaps 2;

```

OY 184 TCATTGTGAGCGATATATGATGAGGACCGCTATGTTGAAAAAGTAAACGAGCAAAAAGCG 243
DB 238 TCAATTGCGCGATATATCCTTGAAGTCCCTATATTCGCGAAATCGGTGAAGAGATGCT 297
OY 244 GCAGACAGTATTTTCAATATGTTGAACCGATCTTGAAGCTCGGATTATGACAGCA 303
DB 298 TAGAGGAATTTTTCATATATGTTAAACCAATTTGACGATGAGACTTGGCTGGCC 357
OY 304 AACTTGTAAAAACCGGTAACTTCAAAAAGATTATA-----ACAAGA 348
DB 358 AATCTTGAATCTGCCCTCACTATGACAAATCGAACTTACGTTAAGCCGCTCAAGTTTA 417
OY 349 GATTAAGAGATTCATCGACAGACGATTAAGAAATCAGTAAGTCTTGAAGATATGAAT 408
DB 418 AGTGAAGATTTTATTTGATTTCTTCATTTGACTGACGTAAGCAAGAGTGGC 477
OY 409 TTCACGGTTCTCAACAGCGCAACAACAAGATGATTAACGCGCTTACGCGCATGAAA 468
DB 478 ATTAGCCTGATTTCCATGCGCAACAACAATACCGGCAACATGCGGTAATGTT 537
OY 469 GATACCTTGAGAAATTTGCGAAGCAAACTTGATATCGTTGACCGGATACAGCTTA 528
DB 538 GACGCGATGAAATTTAAGAGATGAGAAATGATTAATTTGAAATGGCCATGATCGA 597
OY 529 AGTGATGCGAAAAAATTTGCTACAGAAAGTCAACGGGGTAAACGATTGCAACGCTT 588
DB 598 GCAGAAAGC--TGCCAAAGCTTATCAATTTGTCGATATATTTGACTACCTCATTTT 654
OY 589 GCGTTTACCGATGTGTCGCGGAAAGTTTTCGCGCTTAAAAAGATACGCGGCGTGTG 648
DB 655 GCAGTATGAGATGTGATTAACCTGGCGCAAGCAGTATGATGATGAACAGAGTATCTC 714
OY 649 CCGCGAGATCTGAATCTTCAATCCTATGATTTCAAGACGAAACAAACATGCTGACAT 708
DB 715 ACAGACCAATCCCAAGGCTTTCTTAACCTGCAATAGCTATGACCAAGATCTGATTTG 774
OY 709 GTTGTGTGAGTCACACCTGGGCGCAAGATATGACAAATGATCAACGCGCAGCGC 768
DB 775 GTTATGCGCTATATTCATGCTGGAATGATATATTTGTCACCTGATGCAACACACAG 834
OY 769 CAGCTTGCAAGACCATGTCTGATGCGGAGCTGACATCATTCGTGCGCATCATCCGAC 828
DB 835 GAGCTGCTGAAGACCTTAATTTGATGCTGATGAGATTTGTTATTTGCTCCACACCCAC 894
OY 829 GTCTTGAACCGATTTGAATATATACGAAACGCTCATTTTCTACAGCTTCGCAACTT 888
DB 895 TCCCTCTGCTGTTGAAAAATACCAAGATGGATTTATTTTATATGCTTGGGCAATTTT 954
OY 889 GTCTTGAACCAAGCTGAGAGAAACAAGACATGCACTGCTGATCATCCTGAAG 948
DB 955 ATTTTGACCAAGGATGCAAAAGTTCAACCGATTCGTTATTTTGAACATGATCATAT 1014

```

```

OY 949 AAAATGCAACAGCGCGCTTGAAGTACACCGATCATATTCATGAACGACACTGCA 1008
DB 1015 AGTCCAGACCGAGTTTCGTTTACCTTAAGGCCAATGAAAAATAGAAAGCTATTCCCAA 1074
OY 1009 CC 1010
DB 1075 CC 1076

RESULT 10
ADB06119
ID ADB06119 standard; DNA; 1212 BP.
XX
AC ADB06119;
XX
DE 20-NOV-2003 (first entry)
XX
DE Allotiococcus ostitis antigenic protein encoding DNA SEQ ID NO:59.
XX
KW Allotiococcus ostitidis; antigenic protein; immunogenic; immunisation;
XX gene therapy; Gram-positive bacterium; infection; gene; de.
XX
OS Allotiococcus ostitis.
XX
PN MO2003048304-A2.
XX
PD 12-JUN-2003.
XX
PF 25-NOV-2002; 2002WO-US036123.
XX
PR 29-NOV-2001; 2001US-0333777P.
XX PR 18-NOV-2002; 2002US-0426742P.
XX
PA (AMMP) WYETH HOLDINGS CORP.
XX
PI Fletcher LD, McMichael JC, Russell DF, Zagursky RJ;
XX
DR WPI; 2003-505284/47.
XX
DR P-PSDB; ADB06120.
XX
PT New Allotiococcus ostitidis polynucleotides and polypeptides, useful for
XX treating and diagnosing diseases, drug screening assays and monitoring of
XX effects during drug clinical trials.
XX
PS Claim 7; SEQ ID NO 59; 1019pp; English.
XX
XX
XX The present invention describes an isolated polynucleotide (1) of
XX Allotiococcus ostitidis genomic DNA, which encodes an antigenic protein.
XX Allotiococcus ostitidis is a Gram-positive bacterium. Also described: (1)
XX an isolated polypeptide that is encoded by the polynucleotide (1); (2) an
XX expression vector comprising the novel isolated polynucleotide (1), its
XX complement, degenerate variant or fragment; (3) a genetically engineered
XX host cell, transfected, transformed or infected with the vector of (2);
XX (4) an antibody specific for the polypeptide of (1); (5) an immunogenic
XX composition comprising the polypeptide, its complement, biological
XX equivalent or fragment, or the polynucleotide that is comprised in the
XX expression vector; (6) a pharmaceutical composition comprising the
XX polypeptide of (1) and a carrier; (7) a protein chip comprising an array
XX of the polypeptides of (1), their biological equivalent or fragment; (8)
XX immunising against Allotiococcus ostitidis by administering to a host the
XX immunogenic composition; (9) detecting and/or identifying Allotiococcus
XX ostitidis in the biological sample; (10) a kit comprising a container
XX containing the novel polynucleotide, its degenerate variant or fragment,
XX or the antibody of (4); and (11) producing a polypeptide by culturing the
XX genetically engineered host cell under conditions suitable to produce the
XX polynucleotide from the culture. (1) can be used in gene therapy. The
XX polynucleotides, polypeptides, antibodies and compositions of the present
XX invention can be used for treating and diagnosing diseases, drug
XX screening assays and monitoring of effects during drug clinical trials.
XX The polynucleotides are useful for expressing and detecting Allotiococcus
XX ostitidis. The present sequence encodes an Allotiococcus ostitidis antigen
XX protein from the present invention.
XX

```

Sequence 1212 BP; 353 A; 248 C; 259 G; 352 T; 0 U; 0 Other;

Query Match	Score	DB	Length
8.5%	96.4	9	1212

Best Local Similarity 46.7%; Pred. No. 7.8e-17;
Matches 203; Conservativity 0; Mismatches 431

Matches	393;	Conservative	0;	Mismatches	431;	Indels	18;	Gaps	27;
---------	------	--------------	----	------------	------	--------	-----	------	-----

Qy	184	TCATTTGTAGCGCATTTATGATAGGAGCGTATGTTGAAAAAGTAAACGAGCAAAAAAGG	243
Db	238	TCATTTGTGGCGCATATATCCCTTAGTCGCTATATTTGGGAAATCGGTGAACGAGATGGT	297
Qy	244	GCACAGCATATTTTTCATATGTTGAACCGCATCTTATAGCGCTGGATATATGTACAGGA	303
Db	298	TACGAGGATTTTTCATATGTTTAAACACATTTTTCAGGTAGAGATGGTGGTGGCC	357
Qy	304	AACTTTGAAAAACCGGTAACTTATCAAAAGATTATTA-----ACAAACA	348
Db	358	AATCTTGAATGTGCGGTCACTATGACAAATGAACTTACGTTAAGCCGGCTTCAAGTTTA	417
Qy	349	GATTAAGAGATTCAATGTCAGACGATATAGAAATCAGTAAAGTCTTGAAGATATGAAAT	408
Db	418	AGTGAGAGTATTTTATTGGATTCTTCCATTTGACTGACGTAAGGACATGCAAGAAAGCTGGC	477
Qy	409	TTCAACGGTTCACAAAGGCCAACCAACCCAGCATGATTTACGGGTTCAGGAGATGAAA	468
Db	478	ATTGACCTAGTTTCCATGCGCCAAACACATACCGCGCATGCGGCAAGCAGAGTATGGTT	537
Qy	469	GATACGCTTGGAGAAATTTGCGAAGCAAAACCTTGATATCGTTGAGCGGATACAGCTTA	528
Db	538	GACGGCATGGAATTTTAAAGATATGTAAATGATTTATTTGGAATGGCGCATGATGCA	597
Qy	529	AGTATGCGAAAAAGAAATTTTCGTACCAAGAAAGTCAAAGGGAGTACGATTGCAACGCTT	588
Db	598	GCAGAAAGC---TGCCCAAGCTTATCAATTTTGTGGAAATTAATTTGACTACATTCATTTTT	654
Qy	589	GCGTTTACCGATGTGTCCGGGAAAGGTTTCGCGGTAAAGAAATACGCCCGGCGTGTG	648
Db	655	GCAGTTAGCGATGATTTAAACCTGGCCMAAGCTAGTGAATGAACAAGGTACTTC	714
Qy	649	CCCGCAGATCTGAAATCTTCATCCCTATGATTTTCAGAAAGGAAAAACATGCTGCACAT	708
Db	715	ACGACCAACTCCAGACCTTTCTTAACTCTGCGCAATGACTATGACCAAGAAATCTGATTTG	774
Qy	709	GTTGTTGTGCACTACACTCGGGGCCAAGATATGACATGATCCAAAGCAACCGCCAGCGC	768
Db	775	GTTATTCGCTATATCCATGCTGGAATATAGTATATTCGTCAACCTGATGCAACCAACAG	834
Qy	769	CAGCTTGCAAGAGCCATGTCTGATGCGGGAAGTGCATATCGTGGCCATCATCCGAC	828
Db	835	GAGCTGGCTGAAAGCTTATATGATGCTGTGCAGATATGTATTTGCTCCACACCCAC	894
Qy	829	GTCATTGAACCGATTGAGTATATTAAGGAAACCGCATTTTCTACAGCTCGGCAACTT	888
Db	895	TCCCTCCTGCTGTGTAAGAAATATCCMAATGGGATTAATTTTATGAGCTTGGCAATTTT	954
Qy	889	GTCATTGACCAAGGCTGAGCAGAAACAAGACAGTGCATGCTTCACTATACCTGAAG	948
Db	955	ATTTTGGCAAGGATGCAAAAGTTCAACCGATTTCCGTTATTTTGAATGATGATATCAAT	1014
Qy	949	AAAAATGACAGGCGCTTGAAGTGAACCCGATGATATTCATGAAGCAGACACTGCA	1008
Db	1015	AGTCACAGACAGGTTCTGTTTAACTTTAAGGCCAATGAATAAGAAAGCTGATTTCCAAA	1074
Qy	1009	CC 1010	
Db	1075	CC 1076	

RESULT 11

ID ADB06117 standard; DNA; 1212 BP.

ADB06117;

```

DT      20-NOV-2003      (first entry)
XX      Allotiococcus ocltis antigenic protein encoding DNA SEQ ID NO:57.
DE
XX      Allotiococcus ocltis; antigenic protein; immunogenic; immunisation;
XX      Allotiococcus ocltis; antigenic protein; immunogenic; immunisation;
XX      gene therapy; Gram-positive bacterium; infection; gene; ds.
XX      Allotiococcus ocltis.
OS      Allotiococcus ocltis.
PN      WO2003048304-A2.
PD      12-JUN-2003.
PE      25-NOV-2002; 2002WO-US036123.
PR      29-NOV-2001; 2001US-0333777P.
PR      18-NOV-2002; 2002US-0426742P.
XX      (AMHP ) WYETH HOLDINGS CORP.
PA
XX      Fletcher LD, Mcmichael JC, Russell DP, Zagursky RJ,
PI      WPI; 2003-505284/47.
DR      P-PSDB; ADB06118.
XX
XX      New Allotiococcus ocltis polynucleotides and polypeptides, useful for
PT      treating and diagnosing diseases, drug screening assays and monitoring of
PT      effects during drug clinical trials.
XX
XX      Claim 7; SEQ ID NO 57; 1019pp; English.
PS
XX      The present invention describes an isolated polynucleotide (I) of
CC      Allotiococcus ocltis genomic DNA, which encodes an antigenic protein.
CC      Allotiococcus ocltis is a Gram-positive bacterium. Also described: (1)
CC      Allotiococcus ocltis is a Gram-positive bacterium. Also described: (1)
CC      an isolated polypeptide that is encoded by the polynucleotide (I); (2) an
CC      expression vector comprising the novel isolated polynucleotide (I), its
CC      complement, degenerate variant or fragment; (3) a genetically engineered
CC      host cell, transfected, transformed or infected with the vector of (2);
CC      (4) an antibody specific for the polypeptide of (1); (5) an immunogenic
CC      composition comprising the polypeptide, its complement, biological
CC      equivalent or fragment, or the polynucleotide that is comprised in the
CC      expression vector; (6) a pharmaceutical composition comprising the
CC      polypeptide of (1) and a carrier; (7) a protein chip comprising an array
CC      of the polypeptides of (1), their biological equivalent or fragment; (8)
CC      immunising against Allotiococcus ocltis by administering to a host the
CC      immunogenic composition; (9) detecting and/or identifying Allotiococcus
CC      ocltis in the biological sample; (10) a kit comprising a container
CC      containing the novel polynucleotide, its degenerate variant or fragment,
CC      or the antibody of (4), and (11) producing a polypeptide by culturing the
CC      genetically engineered host cell under conditions suitable to produce the
CC      polypeptide from the culture. (1) can be used in gene therapy. The
CC      polynucleotides, polypeptides, antibodies and compositions of the present
CC      invention can be used for treating and diagnosing diseases, drug
CC      screening assays and monitoring of effects during drug clinical trials.
CC      The polynucleotides are useful for expressing and detecting Allotiococcus
CC      ocltis. The present sequence encodes an Allotiococcus ocltis antigen
CC      protein from the present invention.
XX
SQ      Sequence 1212 BP; 353 A; 248 C; 259 G; 352 T; 0 U; 0 Other;

Query Match      8.5%; Score 96.4; DB 9; Length 1212;
Beat Local Similarity 46.7%; Pred. No. 7.8e-17;
Matches 393; Conservative 0; Mismatched 431; Indels 18; Gaps 2;

QY      TCATTGTAGAGCGATATTATGATGAGGACGCTATGTTGAAAAAGTAAACGAGCAAAAAGGG 243
Db      TCAATTGTGGGCGATATATCTTGAAGTCGCTATATGCGGAAATCGGGAACGAGATGCT 297
QY      GCAGACAGTATTTTTCATATATGTTGAACCGATCTTGAAGCCGTCGATATATGACAGAGA 303
Db      TACAGAGCAATTTTTCATATATGTTAAACACATTTTGAACGCTGAAGACTTGTCGTTCC 357
QY      304 AACTTTGAAAACCCGGTACCTATCAAAAAGATTATAA-----ACAGCA 348

```

Query Match	8.5%	Score 96.4	DB 9	Length 1212
-------------	------	------------	------	-------------

Best Local Similarity 46.7%; Pred. No. 7.8e-17;

Matches 393; Conservative 0; Mismatches 431; Indels 18; Gaps 2;

QY 184 TCATTGTAGCGGATATTATGATGGACGCTATGTTGMAAAGTACCGACCAAAAAGG 243

Db 238 TCATTGTGGCGGATATATCTTGTAGTCGTATATGTGGGAAATCGGTGAACGAGATG 297

QY 244 GCAGACGATATTTTCAATATGTGAACCGATCTTTAGAGCCTCGGATTATGTAGCAGSA 303

Db 298 TACGAGGAAATTTTACATATGTATTAACACATTTTGCAGGTAGAGACTTGTCGTTGCC 357

QY 304 AACTTGMAAACCCGGTACCTATCAAAAAGATTATA-----ACAGCA 348

```

DB 358 AATCTTGAATCTGCGCTCACTATGCAATGCAACTTACGTTACCGCGCTTCAAGTTTA 417
QY 349 GATTAAGAGATTCATCTGCAGACGAATAGGAATGAGTAAGTCTTGAAGATATGAAAT 408
DB 418 AGTGAAGGATTTATTTGATTTCTTCATGTGACTGACAAAGCATGAGAACCTGGAC 477
QY 409 TTCACGCTTCAACAGCGCCCAACCAACGCAATGATTTACGGGCTTCAGGGCATGAAA 468
DB 478 ATTGACCTAGTTTCCATGGCCCAACCACTACCGCGCATGCGGCAAGAGGATGTTG 537
QY 469 GATAGGCTTGGAGATTTGGCAAGCAAACTTATGCTTGGAGCGGATACAGCTTA 528
DB 538 GACGGCATGGAATTTTAAAGATGTAAGATGATTTATTTGGATGGCCATGATGCA 597
QY 529 AGTATGCGAAAAAGAAATTTCTGACCAAGAAAGTCAACGGGGTACGATTCACGCTT 588
DB 598 GCGAAGAC---TGCCCAAGCTTATCAATTTGTTGGAAATTAATTTGACTACCTCAATTTT 654
QY 589 GGCCTTACCGATGTGTCGGGAAAGTTTCGCGCTTAAAGAAATACGCGCGGCTGCTG 648
DB 655 GCAGTTAGGATGATTAATTAACCTGGCCAAAGCAGTATGATGATGAAACAGGTGATCTC 714
QY 649 CCGCGAGATCCCGAATCTTCACTCCCTATGATTTACAGAGCAAAAAATGATGATCATT 708
DB 715 AGACCAACTCCCAAGCTTTCTTAACTGGCCATAGCTATGACCAAGAAATCTGATTTG 774
QY 709 GTTGTGTGACATGACACTGGGGCCAAAGATGACATGATCCAAACGCGCCAGCC 768
DB 775 GTTATGCGCTATATCTATCTGTAATGATATTTGTCGCAACCTGATCAAAACGACGAG 834
QY 769 CAGCTTCAAGAGCCATGCTGATGCGGAGCTGACATCATGCTCGGCATCATCCGAC 828
DB 835 GAGCTGCTGAAAGCTTAATGATGCTGGTGCAGATATTTGTTATTTGCCACACCCAC 894
QY 829 GCTTATGAACCATTAATATATTAACGAAACCGTATTTTCAAGCCCTCGGCAACTT 888
DB 895 TCCCTCTCCCTGTTTAAATAATCAAGATGGGATTTATTTTATGCTTGGCAATTTT 954
QY 889 GTCCTTGAACAGGCTGAGCAGACAAAGACAGTGCAGTGTGATGATCACTGAAG 948
DB 955 ATTTTGAACCAAGGATGCAAACTTCAACGATTCCTTATTTTAAACATGATATCAAT 1014
QY 949 AAAAATGAACAGCGCTTGAAGTACACCGATGATATCCATGAAGCAGCACTGCA 1008
DB 1015 AGTCCAGACAGGTTGTTTACCTTAAGCCAAATGAAGTGTATTTCCCAAA 1074
QY 1009 CC 1010
DB 1075 CC 1076

```

RESULT 12
ADBI2064_00/c
WP Sequence Split into 18 fragments LOCUS ADBI2064 Accession Adbi2064

Fragment Name	Begin	End
WP ADBI2064_00	1	110000
WP ADBI2064_01	100001	210000
WP ADBI2064_02	200001	310000
WP ADBI2064_03	300001	410000
WP ADBI2064_04	400001	510000
WP ADBI2064_05	500001	610000
WP ADBI2064_06	600001	710000
WP ADBI2064_07	700001	810000
WP ADBI2064_08	800001	910000
WP ADBI2064_09	900001	1010000
WP ADBI2064_10	1000001	1110000
WP ADBI2064_11	1100001	1210000
WP ADBI2064_12	1200001	1310000
WP ADBI2064_13	1300001	1410000
WP ADBI2064_14	1400001	1510000
WP ADBI2064_15	1500001	1610000
WP ADBI2064_16	1600001	1710000

```

WP ADBI2064_17 1700001 1754382
ID ADBI2064 standard; DNA: 1754382 BP.
XX
AC ADBI2064;
XX
DT 20-NOV-2003 (first entry)
XX
DE Alloicoccus ocltis entire genome sequence SEQ ID NO:6651.
XX
KM Alloicoccus ocltis; antigenic protein; immunogenic; immunisation;
XX gene therapy; Gram-positive bacterium; infection; gene; ds.
XX
OS Alloicoccus ocltis.
XX
PN WO2003048304-A2.
XX
PD 12-JUN-2003.
XX
PF 25-NOV-2002; 2002MO-US036123.
XX
PR 29-NOV-2001; 2001US-0333777P.
XX 18-NOV-2002; 2002US-0426742P.
XX
PA (AMHP ) WYETH HOLDINGS CORP.
XX
PI Fletcher LD, McMichael JC, Russell DP, Zagursky RJ;
XX
DR WPI; 2003-505284/47.
XX
PT New Alloicoccus ocltis polynucleotides and polypeptides, useful for
PT treating and diagnosing diseases, drug screening assays and monitoring of
PT effects during drug clinical trials.
XX
PS Example 3; SEQ ID NO 6651; 1019pp; English.
XX
CC The present invention describes an isolated polynucleotide (1) of
CC Alloicoccus ocltis genomic DNA, which encodes an antigenic protein.
CC Alloicoccus ocltis is a Gram-positive bacterium. Also described: (1)
CC an isolated polypeptide that is encoded by the polynucleotide (1); (2) an
CC expression vector comprising the novel isolated polynucleotide (1), its
CC complement, degenerate variant or fragment; (3) a genetically engineered
CC host cell, transfected, transformed or infected with the vector of (2);
CC (4) an antibody specific for the polypeptide of (1); (5) an immunogenic
CC composition comprising the polypeptide, its complement, biological
CC equivalent or fragment, or the polynucleotide that is comprised in the
CC expression vector; (6) a pharmaceutical composition comprising the
CC polypeptide of (1) and a carrier; (7) a protein chip comprising an array
CC of immunising against Alloicoccus ocltis by administering to a host the
CC immunogenic composition; (9) detecting and/or identifying Alloicoccus
CC ocltis in the biological sample; (10) a kit comprising a container
CC containing the novel polynucleotide, its degenerate variant or fragment,
CC or the antibody of (4); and (11) producing a polypeptide by culturing the
CC genetically engineered host cell under conditions suitable to produce the
CC polypeptide from the culture. (1) can be used in gene therapy. The
CC polynucleotide, polypeptides, antibodies and compositions of the present
CC invention can be used for treating and diagnosing diseases, drug
CC screening assays and monitoring of effects during drug clinical trials.
CC The polynucleotides are useful for expressing and detecting Alloicoccus
CC ocltis. The present sequence represents the entire genome of
CC Alloicoccus ocltis, which is given in the exemplification of the
CC present invention.
XX
SQ Sequence 1754382 BP; 484756A; 391265C; 387369G; 490992T; 0U; 00Other;
XX

```

Query Match 8.5%; Score 96.4; DB 9; Length 110000;
Best Local Similarity 46.7%; Pred No. 6 2e-16;
Matches 393; Conservative 0; Mismatches 431; Indels 18; Gaps 2;

```

QY 184 TCATTGTGAGCATATTAATGATGGACCGCTATGTTGAAAAAGTAAAGGAGCAAAAGGG 243
DB 24408 TCATTGTGAGCATATTAATGATGGACCGCTATGTTGAAAAAGTAAAGGAGCAAAAGGATGCT 24349

```

```

Qy 244 GCAGACGATATTTTTCATATGTTGAACCGATCTTTAGAGCTCGGATTATGACAGA 303
Db 24348 TAGACGAGATTTTTCATATGTTGAACCGATCTTTAGAGCTCGGATTATGACAGA 24289
Qy 304 AACTTTGAAAACCGGTAACCTATCAAAAAGATTATA-----ACAGCA 348
Db 24288 AATCTTGAAATGCGCTGACCTATGACAACTTACGTTAAGCCGGCTTCAAGTTTA 24229
Qy 349 GATAAAGATTTATCTGACAGCAATAGGAATAGGAAGTCTTGAAGATATGAT 408
Db 24228 AGTGAGGATATTTATTTGATTTCTTCCATTGACTGACAAAGCCATCAAGAAAGCTGC 24169
Qy 409 TTCAAGGTTCTCAACAGCCCAACCAACCAATGATTAAGCGGCTTCAAGGCAATGAAA 468
Db 24168 ATTGACCTAGTTTCCATATGCGCAACACATACCGGCGCATGAGGCAACGAGATGTTT 24109
Qy 469 GATACGCTTGAGAAATTTGCGAAGCAAACTTGATATCGTTTGAAGCGGATACAGCTTA 528
Db 24108 GACGCGATGAAATTTTAAAGATGTAGAAATGATTAATTTGGAATGGCCCATGATCGA 24049
Qy 529 AGTATGCGAAGAAAAGAAATTTCTACAGAAAGTCAAGCGGGTAAAGATTGCAACGCTT 588
Db 24048 GCAGAAAGC---TGCCAAAGCTTATCAATTTGTTGCAATTAATTGACTACCTCCATTTT 23992
Qy 589 GCGTTTACCGATGTGTCCGGAAGGTTTCGCGGCTAAAGAAATACCGCGGCGCTGCTG 648
Db 23991 GCAGTTAGCGATGTATTAACCTGCGCAAGCAGTATGATGATGAAACGAGTGTACTC 23932
Qy 649 CCGCGAATCTGAAATCTTCATCCCTATGATTTGAGAGCGAAGAAACATGCTGACATT 708
Db 23931 ACGACCACTCCCAAGCCTTTCTTAACCTGGCAATAGCTATGACCAAGAAATCTGATTTG 23872
Qy 709 GTTGTGTGACGTACACTGGGGCCCAAGATATGACAAATGATCCAAAGCCGCAAGGC 768
Db 23871 GTTATGCGCTTATCACTGCTGAATAGATATATTCGTCACCTGATGCAAAACCAAG 23812
Qy 769 CAGCTTGCAAGAGCCATCTGATGCGGAGCTGACATCATCTGCGCATCATCCGAC 828
Db 23811 GAGCTGGCTGAAGAGTTATGATGCTGTGCAATTTGTTATTTGCTCCACAGCCAC 23752
Qy 829 GTCTTGAACCGATGGAATATATACGAAACCGTCACTTTTCTACAGCCTGGCACTTT 888
Db 23751 TCCCTCTGCGCTGTAAGAAATATCCAAAGATGGAATTTATTTATGCGGCAATTTT 23692
Qy 889 GTCTTGAACCAAGGCTGACAGCAAGACAGACATGACGTGCTCAGTATCACTGAAG 948
Db 23691 ATTTTGAACCAAGGATCAAAAGTTCAACCGATTCGTTATTTTGAACATGGAATCAAT 23632
Qy 949 AAAAATGAAACAGCGCTTGAAGTGAACCGATCGATTCATGAAAGCGACACTGCA 1008
Db 23631 AGTCCAGACAGGTTCTTTTACCTTAAGGCCAATGAAGCTGTGATTTCCCAA 23572
Qy 1009 CC 1010
Db 23571 CC 23570

```

RESULT 13
ABQ69245_05/c
Continuation (6 of 31) of ABQ69245 from base 500001 (Listeria innocua contig DNA sequence #684.
WP Sequence split into 31 fragments LOCUS ABQ69245 Accession Abq69245

```

WP Fragment Name Begin End
WP ABQ69245_00 1 110000
WP ABQ69245_01 100001 210000
WP ABQ69245_02 200001 310000
WP ABQ69245_03 300001 410000
WP ABQ69245_04 400001 510000
WP ABQ69245_05 500001 610000
WP ABQ69245_06 600001 710000
WP ABQ69245_07 700001 810000
WP ABQ69245_08 800001 910000
WP ABQ69245_09 900001 1010000
WP ABQ69245_10 1000001 1110000

```

```

WP ABQ69245_11 1100001 1210000
WP ABQ69245_12 1200001 1310000
WP ABQ69245_13 1300001 1410000
WP ABQ69245_14 1400001 1510000
WP ABQ69245_15 1500001 1610000
WP ABQ69245_16 1600001 1710000
WP ABQ69245_17 1700001 1810000
WP ABQ69245_18 1800001 1910000
WP ABQ69245_19 1900001 2010000
WP ABQ69245_20 2000001 2110000
WP ABQ69245_21 2100001 2210000
WP ABQ69245_22 2200001 2310000
WP ABQ69245_23 2300001 2410000
WP ABQ69245_24 2400001 2510000
WP ABQ69245_25 2500001 2610000
WP ABQ69245_26 2600001 2710000
WP ABQ69245_27 2700001 2810000
WP ABQ69245_28 2800001 2910000
WP ABQ69245_29 2900001 3010000
WP ABQ69245_30 3000001 311208

```

Query Match 6.8%; Score 77.8; DB 6; Length 110000;
Best Local Similarity 57.0%; Pred. No. 1.4e-10;
Matches 142; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

```

Qy 649 CCGCGAATCTGAAATCTTCATCCCTATGATTTCAAGAGCAAAAAACATGCTGACATT 708
Db 37024 CCAGCTATCTTGAAGAAATTTGTAAGACGTTAAAAATATTAAGAAAGATACGCTC 36965
Qy 709 GTTGTGTGACGTACACTGGGGCCAAAGATGACATGATCAATGCCAAAGACCGCAGCC 768
Db 36964 GTTATTTGATTAATGACACTGGGGCGCTCAATATCTGAAACTCCAAACTCAACAACT 36905
Qy 769 CAGCTTGCAAGAGCCATCTGATGCGGAGCTGACATCATCTGCGCATCATCCGAC 828
Db 36904 CAATTTGACATGCTATTTTATGATGCTGTGCTGATATTAATTTAGGAGCTCATCCAC 36845
Qy 829 GTCTTGAACCGATGGAATATATACGAAACCGTCACTTTTCTACAGCCTGGCACTTT 888
Db 36844 CGCCTGGAAGATATCGAAAAATATTAAGATTAATATTTGATACAGATGGCGGATTTT 36785
Qy 889 GTCTTTGAC 897
Db 36784 GCTTTTGCC 36776

```

RESULT 14
ABQ67197_04/c
Continuation (5 of 12) of ABQ67197 from base 400001 (Listeria innocua contig DNA sequence
WP Sequence split into 12 fragments LOCUS ABQ67197 Accession Abq67197

```

WP Fragment Name Begin End
WP ABQ67197_00 1 110000
WP ABQ67197_01 100001 210000
WP ABQ67197_02 200001 310000
WP ABQ67197_03 300001 410000
WP ABQ67197_04 400001 510000
WP ABQ67197_05 500001 610000
WP ABQ67197_06 600001 710000
WP ABQ67197_07 700001 810000
WP ABQ67197_08 800001 910000
WP ABQ67197_09 900001 1010000
WP ABQ67197_10 1000001 1110000
WP ABQ67197_11 1100001 1163020

```

Query Match 6.8%; Score 77.8; DB 6; Length 110000;
Best Local Similarity 57.0%; Pred. No. 1.4e-10;
Matches 142; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

```

Qy 649 CCGCGAATCTGAAATCTTCATCCCTATGATTTCAAGAGCAAAAAACATGCTGACATT 708
Db 46758 CCAGCTATCTTGAAGAAATTTGTAAGACGTTAAAAATATTAAGAAAGATACGCTC 46699
Qy 709 GTTGTGTGACGTACACTGGGGCCAAAGATGACATGATCAATGCCAAAGACCGCAGCC 768

```

DB 46698 GTTATTGTTAATATGCACTGGGCGCTGCATATCGTGAACCTCCAAACCACTACCAACT 46639
QY 769 CAGCTTGCAAGAGCCATGCTGATGCGGAGCTGACATCATGCTGGCCATCATCCGAC 828
DB 46638 CATTGTCGACATGCTATTTAGATGCTGCTGATATATATATGAGAGCTCATCCGAC 46579
QY 829 GTCTTAGAACCATTTGAAGATATATACGACACCGTCATTTTCTACAGCTCCGCAACTT 888
DB 46578 CCGCTGGAAGATGCAAAAATATAAAGATATATATGTTATACAGTATGCGCATTTT 46519
QY 889 GTCTTTGAC 897
DB 46518 GCTTTTGGC 46510

RESULT 15

AB068713 standard; DNA; 885 BP.

AC AB068713;

XX 29-AUG-2003 (revised)

DT 29-AUG-2002 (first entry)

XX Listeria monocytogenes 4b contig DNA sequence #1479.

XX Antibacterial; Listeria; food contamination; mutational analysis;

XX infection; ds.

XX Listeria monocytogenes ATCC 19115.

XX WO200228891-A2.

XX 11-APR-2002.

XX 04-OCT-2001; 2001WO-FR03061.

XX 04-OCT-2000; 2000FR-00012697.

XX (INSP) INST PASTEUR.

XX (CNRS) CNRS CENT NAT RECH SCI.

XX Kunst F, Glaser P;

XX WPI; 2002-332479/37.

XX New genomic sequences from Listeria species, useful for detection,
XX treatment and prevention of infection, also related polypeptides,
XX antibodies and modulators.

XX Claim 14; SEQ ID NO 1526; 180bp; French.

XX The present invention relates to nucleic acid sequences (AB067188-
XX AB071212) from Listeria sp. The sequences are useful as probes and
XX primers for identification and/or detection of Listeria (e.g. as
XX contaminants in foods, or mutational analysis) and for analysis of gene
XX expression. Proteins encoded by the nucleic acid sequences can be used to
XX screen for compounds that modulate gene expression, replication and
XX pathogenicity of Listeria (potential therapeutic agents), also for
XX treating infections by Listeria, and are useful as immunogens in anti-
XX Listeria vaccines. Note: The sequence data for this patent did not form
XX part of the printed specification, but was obtained in electronic format
XX directly from WIPO at http://wipo.int/pub/published_pct_sequences. (updated
XX on 29-AUG-2003 to standardise OS field)

XX Sequence 885 BP; 275 A; 183 C; 161 G; 266 T; 0 U; 0 Other;

XX Query March 6.5%; Score 74; DB 6; Length 885;

XX Best Local Similarity 54.4%; Pred. No. 1.9e-10;
XX Matches 149; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

QY 622 GCTAAGAAATAGCGCGGCGTGTGCTGCCGACAGATCTGAATCTTCATCCCTATGATT 681

DB 40 GCAGACAGCCATACACAGCCGATATCCCATGCTGATATTTGAAGAATTGTACGACGCTG 99
QY 682 TCAGAGCGAAGAAAAACATGCTGACATTTGTTGTGACGTACACTGCGGCGCAAGATAT 741
DB 100 AAGACATACAAAAAGAACTCTGTGCTTATCATATATACGACTGCGGCGTGAATAT 159
QY 742 GACATGATCCAAACGACCGCCAGCGCAGCTTGCAAGAGCCATGCTGATGCGGAGCT 801
DB 160 CGCGAAACACCAACGATTTATCAAAACCAATTTGTCATGCGATTTAGATGCTGTGCT 219
QY 802 GACATCATGTCGGCCATCATCCGACGCTTTAGAACCGATTGAAGTATATACGGAAC 861
DB 220 GATATATTTATGGGCTCTCACCCATCATGCGCTAGAAAGCGTTGAAGATATATAAATAA 279
QY 862 GTCAATTTTCTACAGCTCGGCACTTGTCTTTG 895
DB 280 TACATGTTTATAGCATGCGGCGACCTTGTCTTTG 313

Search completed: February 27, 2006, 07:05:22
Job time : 723.889 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model.

Run on: February 27, 2006, 06:34:28 ; Search time 1248.33 Seconds
(without alignments)
7551.777 Million cell updates/sec

Title: US-10-789-164-3
Perfect score: 1140
Sequence: 1 acgaaagaaactgagctt.....gtgcaactaactaa 1140

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Published Applications NA Main:

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1140	100.0	1140	US-10-484-605-3	Sequence 3, Appl1
2	1140	100.0	1140	US-10-789-164-3	Sequence 6, Appl1
3	1140	100.0	6536	US-10-484-605-6	Sequence 1745, Ap
4	332.4	29.2	636	US-09-974-300-1745	Sequence 1777, Ap
5	225.6	19.8	488	US-09-974-300-1777	Sequence 55, Appl
6	96.4	8.5	1212	US-10-501-282-55	Sequence 57, Appl
7	96.4	8.5	1212	US-10-501-282-57	Sequence 59, Appl
8	96.4	8.5	1212	US-10-501-282-59	Sequence 65, Ap
9	96.4	8.5	1754382	US-10-501-282-6651	Sequence 10, Appl
10	77.8	6.8	1163020	US-10-398-221-10	Sequence 2086, Ap
11	77.8	6.8	3011208	US-10-398-221-2058	Sequence 1526, Ap
12	74	6.5	885	US-10-398-221-1526	Sequence 3171, Ap
13	74	6.5	2282	US-10-398-221-3371	Sequence 514, Ap
14	54.2	4.8	1071	US-10-470-0488-514	Sequence 1055, Ap
15	54.2	4.8	1101	US-10-724-972A-1055	Sequence 86, Appl
16	47.6	4.3	1293	US-10-474-792-53	Sequence 866, App
17	47.6	4.2	1170	US-10-724-972A-866	Sequence 1, Appl1
18	47.2	4.1	536165	US-09-939-964-1	Sequence 34615, A
19	41.8	3.7	2225	US-11-097-143-34615	Sequence 474, App
20	41.8	3.7	3256	US-10-104-047-474	Sequence 261, Ap
21	39.6	3.5	1467	US-10-501-282-2261	Sequence 2263, Ap
22	39.6	3.5	1467	US-10-501-282-2263	Sequence 6651, Ap
23	39.6	3.5	1754382	US-10-501-282-6651	

C 24	39	3.4	620	4	US-09-925-065A-700467	Sequence 700467,
C 25	39	3.4	620	4	US-09-925-065A-700468	Sequence 700468,
C 26	39	3.4	1019	3	US-09-981-566A-96	Sequence 96, Appl
C 27	39	3.4	1205	7	US-10-424-599-107806	Sequence 107806,
C 28	39	3.4	1348	6	US-10-017-161-1059	Sequence 1059, Ap
C 29	39	3.4	1348	6	US-10-292-798-901	Sequence 901, App
C 30	38.6	3.4	1921	3	US-09-764-868-161	Sequence 161, App
C 31	38.6	3.4	2789	7	US-10-161-493-29	Sequence 29, Appl
C 32	38.6	3.4	2812	7	US-10-161-493-27	Sequence 27, Appl
C 33	38.4	3.4	831	6	US-10-033-585-6177	Sequence 6177, Ap
C 34	38	3.3	437	9	US-10-915-740A-935	Sequence 935, App
C 35	38	3.3	549	4	US-09-925-065A-55305	Sequence 55305, A
C 36	38	3.3	5609	9	US-10-915-740A-7	Sequence 7, Appl1
C 37	38	3.3	2242716	9	US-10-915-740A-1068	Sequence 1068, Ap
C 38	37.4	3.3	606	9	US-10-617-320-634	Sequence 634, App
C 39	37.2	3.3	510	6	US-10-029-386-6837	Sequence 6837, App
C 40	37.2	3.3	544	7	US-10-021-323-15363	Sequence 15363, A
C 41	37.2	3.3	590	7	US-10-021-323-14869	Sequence 14869, A
C 42	37.2	3.3	814	8	US-10-767-795-5675	Sequence 5675, Ap
C 43	37.2	3.3	2622	8	US-10-425-115-174562	Sequence 174562,
C 44	37.2	3.3	3710	6	US-10-294-025-1036	Sequence 1036, Ap
C 45	37	3.2	1133	7	US-10-424-599-83576	Sequence 83576, A

ALIGNMENTS

```

RESULT 1
US-10-484-605-3
; Sequence 3, Application US/10484605
; Publication No. US20040253704A1
; GENERAL INFORMATION:
; APPLICANT: Sung, Moon-Hee
; APPLICANT: Hong, Seung-Pyo
; APPLICANT: Lee, Jong-Su
; APPLICANT: Jung, Chang-Min
; APPLICANT: Kim, Chul-Joong
; APPLICANT: Soda, Kenji
; APPLICANT: Ashiuchi, Makoto
; TITLE OF INVENTION: SURFACE EXPRESSION VECTORS HAVING pgsbca, THE GENE CODING
; TITLE OF INVENTION: POLY-GAMMA-GLUTAMATE SYNTHETASE, AND A METHOD FOR EXPRESSION OF
; FILE REFERENCE: 4240-101
; CURRENT FILING DATE: 2004-01-20
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: PCT/KR02/01522
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 1140
; TYPE: DNA
; ORGANISM: Bacillus subtilis
US-10-484-605-3
Query Match 100.0%; Score 1140; DB 8; Length 1140;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 ATGAAAAAGAACTGAGCTTTCATGAAAAGCTTAAGCTGACAAACGAAAAAAG 60
DB 1 ATGAAAAAGAACTGAGCTTTCATGAAAAGCTTAAAGCTGACAAACGAAAAAAG 60
QY 61 AAAACCAATAGACGATTTATTTGCAATCCGATCGTTTGTCTTATGTTGCTTTC 120
DB 61 AAAACCAATAGACGATTTATTTGCAATCCGATCGTTTGTCTTATGTTGCTTTC 120
QY 121 ATGTGGCGGGAAGGGAAGCGGAAAGCGGAAAGCTTAATTTCTGACGACGATCTCA 180
DB 121 ATGTGGCGGGAAGGGAAGCGGAAAGCGGAAAGCTTAATTTCTGACGACGATCTCA 180

```

```
181 GCCTCATTTGAGCGCATATATATGATGGAGCGTATGTTGAAAAAGTAAACGAGCAAAAA 240
181 GCTCATTTGATAGGAGATATATATGATGGAGCGTATGTTGAAAAAGTAAACGAGCAAAAA 240
241 GGGGAGAGAGATATTTTCAATATGTTGAGCCGATCTTTAGAGCTCGGATTTATGTACCA 300
241 GGGGAGAGAGATATTTTCAATATGTTGAGCCGATCTTTAGAGCTCGGATTTATGTACCA 300
301 GGAATCTTGAAGAACCGGTAACCTATCAAAAGAAATTTAAACAGCAGATTAAGAGATT 360
301 GGAATCTTGAAGAACCGGTAACCTATCAAAAGAAATTTAAACAGCAGATTAAGAGATT 360
361 CATCTGACAGAGATTAAGAAATCACTGAAAGTCTTGAAGATATGAAATTTCAAGCTTCTC 420
361 CATCTGACAGAGATTAAGAAATCACTGAAAGTCTTGAAGATATGAAATTTCAAGCTTCTC 420
421 AACAGCGCAACACACGCAATGATTAACGCGCTTCAAGGCGATGAAAGATACGCTTGA 480
421 AACAGCGCAACACACGCAATGATTAACGCGCTTCAAGGCGATGAAAGATACGCTTGA 480
481 GAATTTGGAGAACAAACCTTATATCGTTGAGAGGGGATACAGCTTAAGATACGCAAA 540
481 GAATTTGGAGAACAAACCTTATATCGTTGAGAGGGGATACAGCTTAAGATACGCAAA 540
541 AAGAAATTTTCTGACAGAAAGTCAACGAGGTAAGATGCAACGCTTGGCTTACCGAT 600
541 AAGAAATTTTCTGACAGAAAGTCAACGAGGTAAGATGCAACGCTTGGCTTACCGAT 600
601 GTGTCGGGAAAGGTTTGGCGCTTAAAGAAATACCGCGGCGTCTGCGCGCAGATCTT 660
601 GTGTCGGGAAAGGTTTGGCGCTTAAAGAAATACCGCGGCGTCTGCGCGCAGATCTT 660
661 GAAATCTTCAATCCCTATATGTTGAGAGGAAACATGCTGATGTTGTTGTGAG 720
661 GAAATCTTCAATCCCTATATGTTGAGAGGAAACATGCTGATGTTGTTGTGAG 720
721 TCACACTGGGGGCAAGAGATGATGATGATCCAAACGACCGCGCAGCGCTTGAAGA 780
721 TCACACTGGGGGCAAGAGATGATGATGATCCAAACGACCGCGCAGCGCTTGAAGA 780
781 GCCATGTCGATGCGGAGCTGATCATCGTGGCGCATCATCCGACGCTTTGAAGCG 840
781 GCCATGTCGATGCGGAGCTGATCATCGTGGCGCATCATCCGACGCTTTGAAGCG 840
841 ATTGAAGTATATTAAGGAAACCGTCAATTTTCAACGCTCGGCACTTTGTCTTGAACCA 900
841 ATTGAAGTATATTAAGGAAACCGTCAATTTTCAACGCTCGGCACTTTGTCTTGAACCA 900
901 GGGTGAAGAGAAACAGAGACAGTCACTGTTCACTGATCACTGAAGAAATGAACA 960
901 GGGTGAAGAGAAACAGAGACAGTCACTGTTCACTGATCACTGAAGAAATGAACA 960
961 GGGTGAAGAGAAACAGAGACAGTCACTGTTCACTGATCACTGAAGAAATGAACA 960
961 GGGTGAAGAGAAACAGAGACAGTCACTGTTCACTGATCACTGAAGAAATGAACA 960
1021 GACAGCTTTAAACAGAAACCTTATATGCGGAACTGAGAAAGATCTTATTTCCCTTGG 1080
1021 GACAGCTTTAAACAGAAACCTTATATGCGGAACTGAGAAAGATCTTATTTCCCTTGG 1080
1081 AAGAGTAAAGAGAGAAACCTGAGTTGATGATGATGATGATGATGATGATGATGATGAT 1140
1081 AAGAGTAAAGAGAGAAACCTGAGTTGATGATGATGATGATGATGATGATGATGATGAT 1140
```

RESULT 2
US-10-789-164-3

; Sequence 3, Application US/10789164
; Publication No. US20050191720A1

; GENERAL INFORMATION:
; APPLICANT: Sung, M.H. et al.

; TITLE OF INVENTION: SURFACE EXPRESSION METHOD OF PEPTIDES P5 AND ANAL3 USING THE GENE
; TITLE OF INVENTION: ENCODING POLY-GAMMA-GLUTAMATE SYNTHETASE

```
FILE REFERENCE: P1574  
CURRENT APPLICATION NUMBER: US/10/789,164  
CURRENT FILING DATE: 2004-02-27  
NUMBER OF SEQ ID NOS: 9  
SOFTWARE: KopatentIn 1.71  
SEQ ID NO 3  
LENGTH: 1140  
TYPE: DNA  
ORGANISM: Bacillus subtilis  
US-10-789-164-3
```

Query Match 100.0%; Score 1140; DB 9; Length 1140;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
1 ATGAAAAAGAACTGACCTTTCATGAAAGCTGTAAAGCTGAACAAACAGCAAAAAAG 60
1 ATGAAAAAGAACTGACCTTTCATGAAAGCTGTAAAGCTGAACAAACAGCAAAAAAG 60
61 AAAACCAATAGACAGTATTTATGATGATGATGATGATGATGATGATGATGATGAT 120
61 AAAACCAATAGACAGTATTTATGATGATGATGATGATGATGATGATGATGATGAT 120
121 ATGAGGCGGGAAGAGCGGAAACCGCAAGCTCAAAAGTATTTCTGACGATCTCTCA 180
121 ATGAGGCGGGAAGAGCGGAAACCGCAAGCTCAAAAGTATTTCTGACGATCTCTCA 180
181 GCCTCATTTGAGCGCATATATATGATGGAGCGTATGTTGAAAAAGTAAACGAGCAAAAA 240
181 GCCTCATTTGAGCGCATATATATGATGGAGCGTATGTTGAAAAAGTAAACGAGCAAAAA 240
241 GGGGAGAGAGATATTTTCAATATGTTGAAACCGATCTTGAAGCGCTGATATGATGCA 300
241 GGGGAGAGAGATATTTTCAATATGTTGAAACCGATCTTGAAGCGCTGATATGATGCA 300
301 GGAATCTTGAAGAACCGGTAACCTATCAAAAGAAATTTAAACAGCAGATTAAGAGATT 360
301 GGAATCTTGAAGAACCGGTAACCTATCAAAAGAAATTTAAACAGCAGATTAAGAGATT 360
361 CATCTGACAGAGATTAAGAAATCACTGAAAGTCTTGAAGAGTATGAAATTTCAAGGTTCTC 420
361 CATCTGACAGAGATTAAGAAATCACTGAAAGTCTTGAAGAGTATGAAATTTCAAGGTTCTC 420
421 GGGTGAAGAGAAACAGAGACAGTCACTGTTCACTGATCACTGAAGAAATGAACA 480
421 GGGTGAAGAGAAACAGAGACAGTCACTGTTCACTGATCACTGAAGAAATGAACA 480
481 GAATTTGGAGAAACCTTGAATATGTTGAGAGCGGATACAGCTTAAAGTATGATGCAAA 540
481 GAATTTGGAGAAACCTTGAATATGTTGAGAGCGGATACAGCTTAAAGTATGATGCAAA 540
541 AAGAAATTTTCTGACAGAAAGTCAACGAGGTAAGATGCAACGCTTGGCTTACCGAT 600
541 AAGAAATTTTCTGACAGAAAGTCAACGAGGTAAGATGCAACGCTTGGCTTACCGAT 600
601 GGTGTCGGGAAAGGTTTGGCGCTTAAAGAAATACCGCGGCGTCTGCGCGCAGATCTT 660
601 GGTGTCGGGAAAGGTTTGGCGCTTAAAGAAATACCGCGGCGTCTGCGCGCAGATCTT 660
661 GAAATCTTCAATCCCTATATGTTGAGAGGAAACATGCTGATGTTGTTGTGAG 720
661 GAAATCTTCAATCCCTATATGTTGAGAGGAAACATGCTGATGTTGTTGTGAG 720
721 TCACACTGGGGGCAAGAGATGATGATGATCCAAACGACCGCGCAGCGCTTGAAGA 780
721 TCACACTGGGGGCAAGAGATGATGATGATCCAAACGACCGCGCAGCGCTTGAAGA 780
781 GCCATGTCGATGCGGAGCTGATCATCGTGGCGCATCATCCGACGCTTTGAAGCG 840
781 GCCATGTCGATGCGGAGCTGATCATCGTGGCGCATCATCCGACGCTTTGAAGCG 840
841 ATTGAAGTATTAAGAGAAACCGTCAATTTTCAACGCTCGGCACTTTGTCTTGAACCA 900
841 ATTGAAGTATTAAGAGAAACCGTCAATTTTCAACGCTCGGCACTTTGTCTTGAACCA 900
```


Db 841 ATTGAAGTATATACGGAACCGTCATTTTCTACAGCCTCGGAACTTTGTCTTGACCA 900
Qy 901 GGCTGACGAGAAAGAGACAGTGCCTGTTGATACCTTAAGAAAATGGAACA 960
Db 901 GGCTGACGAGAAAGAGACAGTGCCTGTTGATACCTTAAGAAAATGGAACA 960
Qy 961 GGCGCTTTGAAGTACACCGATTCATATCCATGAAAGCAGACCTGCTGTAAGAAA 1020
Db 961 GGCGCTTTGAAGTACACCGATTCATATCCATGAAAGCAGACCTGCTGTAAGAAA 1020
Qy 1021 GACAGCCTTAACAGAAAACATTAATTCGCAAGTACGAAAGACTTAATTCCTTGG 1080
Db 1021 GACAGCCTTAACAGAAAACATTAATTCGCAAGTACGAAAGACTTAATTCCTTGG 1080
Qy 1081 AAGTAGAAGACGGAACCTGACGTTTGATTTGATGATGACAACTAAATCTTAA 1140
Db 1081 AAGTAGAAGACGGAACCTGACGTTTGATTTGATGATGACAACTAAATCTTAA 1140

RESULT 3

US-10-484-605-6

Sequence 6, Application US/10484605
Publication No. US20040253704A1

GENERAL INFORMATION:

APPLICANT: Sung, Moon-Hee
APPLICANT: Hong, Seung-Pyo
APPLICANT: Lee, Jong-Su
APPLICANT: Jung, Chang-Min
APPLICANT: Kim, Chul-Joong
APPLICANT: Soda, Kenji
APPLICANT: Ashiuchi, Makoto
TITLE OF INVENTION: SURFACE EXPRESSION VECTORS HAVING pgsBCA, THE GENE CODING
TITLE OF INVENTION: POLY-GAMMA-GLUTAMATE SYNTHETASE, AND A METHOD FOR EXPRESSION OF
TITLE OF INVENTION: TARGET PROTEIN AT THE SURFACE OF MICROORGANISM USING THE VECTOR
FILE REFERENCE: 4240-101
CURRENT APPLICATION NUMBER: US/10/484,605
PRIOR FILING DATE: 2004-01-20
PRIOR APPLICATION NUMBER: PCT/KR02/01522
PRIOR FILING DATE: 2002-08-09
PRIOR APPLICATION NUMBER: KR 2001-48373
PRIOR FILING DATE: 2001-08-10
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn version 3.2
SEQ ID NO 6
LENGTH: 6536
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic Construct
US-10-484-605-6

Query Match 100.0%; Score 1140; DB 8; Length 6536;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGAAAAAAGACTGAGCTTTCATGAAAAGCTGCTAAAGCTGACAAAACAGCAAAAAAG 60
Db 1891 ATGAAAAAAGACTGAGCTTTCATGAAAAGCTGCTAAAGCTGACAAAACAGCAAAAAAG 1950
Qy 61 AAAACCAATAGACGATTAATTAATTCATCCATTCGATTCGTTTGTCTTAATGCTTTC 120
Db 1951 AAAACCAATAGACGATTAATTAATTCATCCATTCGATTCGTTTGTCTTAATGCTTTC 2010
Qy 121 ATGTGGCGGGAAGAGCGGAAGCGGAGGTCAAAAGTATTCGACGAGTACTCTCA 180
Db 2011 ATGTGGCGGGAAGAGCGGAAGCGGAGGTCAAAAGTATTCGACGAGTACTCTCA 2070
Qy 181 GCCTCATTTGTAGCGATATTAATGATGAGCGCTATGTTGAAAAAGTAACGAGCAAAAA 240
Db 2071 GCCTCATTTGTAGCGATATTAATGATGAGCGCTATGTTGAAAAAGTAACGAGCAAAAA 2130
Qy 241 GGGGACAGACGATTTTTCATATATGTTGAACCGATCTTTAGAGCTTGGATATATGCA 300

Db 2131 GGGGACAGACGATTTTTCATATATGTTGAACCGATCTTTAGAGCTTGGATATATGCA 2190
Qy 301 GGAACCTTTGAAACCCGGTAACTATCAAAAGATTAATAACAGCAGATAAGAGATT 360
Db 2191 GGAACCTTTGAAACCCGGTAACTATCAAAAGATTAATAACAGCAGATAAGAGATT 2250
Qy 361 CATCTGACAGCAATTAAGCAATCAAGTAAGTCTTGAAGATATGAATTTACGGTCTC 420
Db 2251 CATCTGACAGCAATTAAGCAATCAAGTAAGTCTTGAAGATATGAATTTACGGTCTC 2310
Qy 421 AACAGCGCAACCAACGACGATGATTAACGGCTTCAGGCGCATGAAGATACGTTTGA 480
Db 2311 AACAGCGCAACCAACGACGATGATTAACGGCTTCAGGCGCATGAAGATACGTTTGA 2370
Qy 481 GAATTTGGAAAGCAAAACCTTGATATTCCTTGAACGGGATACAGTTAAGTATGAGCGAA 540
Db 2371 GAATTTGGAAAGCAAAACCTTGATATTCCTTGAACGGGATACAGTTAAGTATGAGCGAA 2430
Qy 541 AAGAAATTTTGTATCAGAAAGTCAACGGGGTAAAGATTCGACAGCTTGGCTTAAACGAT 600
Db 2431 AAGAAATTTTGTATCAGAAAGTCAACGGGGTAAAGATTCGACAGCTTGGCTTAAACGAT 2490
Qy 601 GTGTCCGGGAAAGGTTTCCGGCTTAAAAAAGATTAACCGCGCGTCTGCGCAGATCTT 660
Db 2491 GTGTCCGGGAAAGGTTTCCGGCTTAAAAAAGATTAACCGCGCGTCTGCGCAGATCTT 2550
Qy 661 GAATTTCTTATCTCTTATGATTTTCAAGACGGAAGAAACATGCTGATGTTGTTGCGAG 720
Db 2551 GAATTTCTTATCTCTTATGATTTTCAAGACGGAAGAAACATGCTGATGTTGTTGCGAG 2610
Qy 721 TCACACTGGGGGCAAGAGATGATGACATGATCCAAACGACCGGCAAGCTTTCAGAA 780
Db 2611 TCACACTGGGGGCAAGAGATGATGATGATCCAAACGACCGGCAAGCTTTCAGAA 2670
Qy 781 GCCATGCTGATGCGGAGCTGACATCATGCTGCGCATCATGCGACGCTTTAGAACCG 840
Db 2671 GCCATGCTGATGCGGAGCTGACATCATGCTGCGCATCATGCGACGCTTTAGAACCG 2730
Qy 841 ATTGAAGTATATACGGAACCGTCAATTTTCTACAGCTTCGCAACTTTGTCTTTGACCA 900
Db 2731 ATTGAAGTATATACGGAACCGTCAATTTTCTACAGCTTCGCAACTTTGTCTTTGACCA 2790
Qy 901 GGCTGACGAGAAAGAGAGACAGTGCATGCTGCTAGTATCACCTGAAGAAAATGGAACA 960
Db 2791 GGCTGACGAGAAAGAGAGACAGTGCATGCTGCTAGTATCACCTGAAGAAAATGGAACA 2850
Qy 961 GGCGCTTTGAAGTACACCGATTCATATCCATGAAAGCAGACCTGCACTGTGAAAAAA 1020
Db 2851 GGCGCTTTGAAGTACACCGATTCATATCCATGAAAGCAGACCTGCACTGTGAAAAAA 2910
Qy 1021 GACAGCCTTAACAGAAAACATTAATTCGCAAGTACGAAAGACTTAATTCCTTGG 1080
Db 2911 GACAGCCTTAACAGAAAACATTAATTCGCAAGTACGAAAGACTTAATTCCTTGG 2970
Qy 1081 AAGTAGAAGACGGAACCTGACGTTTGATTTGATGATGACAACTAAATCTTAA 1140
Db 2971 AAGTAGAAGACGGAACCTGACGTTTGATTTGATGATGACAACTAAATCTTAA 3030

RESULT 4

US-09-974-300-1745

Sequence 1745, Application US/09974300

Patent No. US20020146721A1

GENERAL INFORMATION:

APPLICANT: Berka, Randy M.
APPLICANT: Clausen, Ib Groth
TITLE OF INVENTION: Methods For Monitoring Multiple Gene
FILE REFERENCE: 10085,500-US
CURRENT APPLICATION NUMBER: US/09/974,300
PRIOR FILING DATE: 2001-10-05
PRIOR APPLICATION NUMBER: 09/680,598
PRIOR FILING DATE: 2000-10-06

PRIOR APPLICATION NUMBER: 60/279,526
PRIOR FILING DATE: 2001-03-27
NUMBER OF SEQ ID NOS: 8481
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1745
LENGTH: 696
TYPE: DNA
ORGANISM: Bacillus licheniformis
US-09-974-300-1745

Query Match 29.2%; Score 332.4; DB 3; Length 696;
Best Local Similarity 68.4%; Pred. No. 1.1e-85;
Matches 475; Conservative 0; Mismatches 216; Indels 3; Gaps 1;

OY 444 GGATTAGCGGCTTCAGGCGATGAAGATACGCTTGAGATTTGGGAGCAAACTTGA 503
DB 1 GACTACGCGGAGACGGTTTGAAGATACGCTCAATTAATTTTCAATGGAATCTGA 60
OY 504 TATCGTTGAGCGGGATACAGCTTAAGTATGCGAAAAAATTTGTAACAGAAAGT 563
DB 61 GCTTGTGGAGAGGAAATTAATCTTGAAGACGCGAAACGATCTATCAGAAATGT 120
OY 564 CAACGGGTTAAGATTGCAAGCGCTTTCATCCGATGTGTCCGGAAGAGTTTCCGCGC 623
DB 121 GAACGGCGTAAAAATTGCAACGCTCGTTTACAGACGCTTACAAAGAACTTTACAGC 180
OY 624 TAAAAAGAAATCGCGGCGCTGCTCCCGCAGATCTGAATCTTCAATCCCTATGATTTC 683
DB 181 CAAAAAGAACAGAGCGGAGTGTGCGCTCGATCGAAATCTTTATTCCAATGATTGC 240
OY 684 AGAAGCGAAAAACATGCTGACATTTGTTGTGCACTCACATTTGGGCGCAAGATATGA 743
DB 241 GGAAGCATCGAAAAAGCGGATCTTGTCTTGTCCATGTCACCTGGGCAAGAAATATGA 300
OY 744 CAATGATCAAAACGACCGCCAGCGCCAGCTTGCAAGACCATGTGTGATGCGGAGCTGA 803
DB 301 CAATGAACCGAACGACAGACAGAAAGATCTGGCCAAAGCGATTCGAGATGCGGAGCAGA 360
OY 804 CATGATCGTGGCGCATCATCCGCAAGTCTTGAACCGATGAATTAACGGAACCGT 863
DB 361 TGTCAATCATCGCGCTCATCTCCCAATGTTCTCGAACCGATCGAAAGTATACGATACGT 420
OY 864 CATTTTCTACAGCCTCGGCAATTTGTCTTTGACCAAGGCTGGAAGCAACAGAGACAG 923
DB 421 GATTTTCTACAGCCTCGGCAATTTGTATTTGATCAGGCTGTGTCAAGAACCGGAGACAG 480
OY 924 TGCATCTGTTCAATACCTGGAAGAAATGAAACAGCGCGCTTTGAAATGACACCGAT 983
DB 481 CGCGCTTGTACAAATCAATTTAATGAATGACGCAAGGCGCTTTGAGGTAAACGCTCT 540
OY 984 CGATATCATGAAGGACACCTGCACCT--GTGAAAAAAGACAGCCTTAACAGAAAC 1040
DB 541 CAACATTCGCGAAGACGCGCAGCCTTTAGCAAGCGCACTTTTAAACGAAAGC 600
OY 1041 CATTATTCGCGAAGTACGAAAGACTTAATTTCCCTTTGAAAGTGAAGACGGAATCT 1100
DB 601 GATCTTCGCTCAATTTGCAAAAAGAAACAACTCGACTGGAAGAAAGAACGAAATTT 660
OY 1101 GACGTTGATTTGATCATAGTGAACAACTAAA 1134
DB 661 AACGTTGAAGTCATCATGCGACAAAGCTGAA 694

RESULT 5
US-09-974-300-1777
Sequence 1777, Application US/09974300
Patent No. US20020146721A1
GENERAL INFORMATION:
APPLICANT: Berka, Randy M.
APPLICANT: Clausen, Ib Groth
TITLE OF INVENTION: Methods For Monitoring Multiple Gene
FILE REFERENCE: 10085,500-US

CURRENT APPLICATION NUMBER: US/09/974,300
CURRENT FILING DATE: 2001-10-05
PRIOR APPLICATION NUMBER: 09/680,598
PRIOR FILING DATE: 2000-10-06
PRIOR APPLICATION NUMBER: 60/279,526
PRIOR FILING DATE: 2001-03-27
NUMBER OF SEQ ID NOS: 8481
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1777
LENGTH: 488
TYPE: DNA
ORGANISM: Bacillus licheniformis
US-09-974-300-1777

Query Match 19.8%; Score 225.6; DB 3; Length 488;
Best Local Similarity 67.8%; Pred. No. 1.1e-54;
Matches 331; Conservative 0; Mismatches 154; Indels 3; Gaps 1;

OY 1 ATGAAAAAGAACTGACCTTTTCAATGAAGCTGCTAAAGCTGACAAACAGCAAAAAAG 60
DB 1 ATGAAAAAGAACTGACCTTTTCAAGAAAACTGCTGAAGTTGACGAAGCAGAGAAAAAG 60
OY 61 AAAACCAATAGACAGTATTTATTTCCATTCGATCCGATCTTTTGTCTTATGCTTTC 120
DB 61 AAAACCAATAGACAGTATTTATTTGATTTGATTCGCGTATTTTGTCTTAAATGTTGCTTT 120
OY 121 ATGAGGCGGGAAAAAGCGGAAAGC---CGAAGGTCAAAACGTATTCGACGACGTAATC 177
DB 121 ACTTGGGTGGAGAGCGGCAAAATCTTTCGCAATGACAAAAAGAAAGATGCAAGCTT 180
OY 178 TCAACCTATTTTGTAGCGATATTTATGATGAGACGCTATGTTGAAAAAGTACGAGCA 237
DB 181 ACAGTACTTTTGTGGGATATCATATGAGGAAGAAACGTAAGAAAGTACAACTTG 240
OY 238 AAAGGCGACAGATATTTTCAATATGTTGAACCGATCTTTAGACCTCGGATTAATGA 297
DB 241 CACGTTTCGAAAGTGTCTTCAAAATGTGAAGCGGATCTTAATGTGTCAATTTATTC 300
OY 298 GCAGAAACCTTTGAAACCGGTAACCTATCAAAAGATTAATAACAGCAGATAAGAG 357
DB 301 ACAGAAACCTTTGAAACCGGTAACCTATCAAAAGATTAATAACAGCAGATAAGAG 360
OY 358 ATTGATTCGACAGCAATTAAGAAATCAATGAAAGCTTTGAAAGATTAATTTCAAGGTT 417
DB 361 ATCATCTGCAAGCAATCAAGATCAATGCAAACTTTGAAAAAGCTTGAAGCTTCAAGGTA 420
OY 418 CTCAACAGCGCAACCAACGCAATGATTAAGCGCTTCAAGGCAATGAAGATACGCTT 477
DB 421 CTGAATTTTGGCAACCAACATGCGATGAGACTTACGCGGAAAGCGGTTTGAAGGATACGCTC 480
OY 478 GGAGAAAT 485
DB 481 AATGAGTT 488

RESULT 6
US-10-501-282-55
Sequence 55, Application US/10501282
Publication No. US20050203280A1
GENERAL INFORMATION:
APPLICANT: MCMICHAEL, JOHN CALHOUN
APPLICANT: ZAGORSKY, ROBERT JOHN
APPLICANT: RUSSELL, DAVID PARRISH
APPLICANT: FLETCHER, LEAH DIANE
TITLE OF INVENTION: ALLOTIOCCUS OTTIDIS OPEN READING FRAMES (ORFs) ENCODING
FILE REFERENCE: AM100780 L2
CURRENT APPLICATION NUMBER: US/10/501,282
PRIOR APPLICATION NUMBER: 60/333,777
PRIOR FILING DATE: 2001-11-29
PRIOR APPLICATION NUMBER: 60/426,742
PRIOR FILING DATE: 2002-11-18

PRIOR APPLICATION NUMBER: PCT/US02/36123
PRIOR FILING DATE: 2002-11-25
NUMBER OF SEQ ID NOS: 6653
SOFTWARE: PatentIn version 3.2
SEQ ID NO 55
LENGTH: 1212
TYPE: DNA
ORGANISM: Alloiococcus otitidis
FEATURE:
NAME/KEY: CDS
LOCATION: (463) ..(1209)
US-10-501-282-55

Query Match 8.5%; Score 96.4; DB 9; Length 1212;
Best Local Similarity 46.7%; Pred. No. 7, 7e-17;
Matches 393; Conservative 0; Mismatches 431; Indels 18; Gaps 2;

184 TCATTGTAGGGGATATATGATGGAGCGCTATGTTGAAAAGTAAACGAGCAAAAAGG 243
238 TCATTGTAGGGGATATATGATGGAGCGCTATGTTGAAAAGTAAACGAGCAAAAAGG 297
244 GCAGACAGTATTTTCAATATGTTGAACGATCTTGAAGCTCGGATTATGACAGA 303
298 TAGAGAGATTTTTCATATGTTAAACCAATTTGACGGTAGAGACTTGGCTGGCC 357
304 AACTTGAACCCGGTAACTATCAAAAAGATTATA-----ACAAGCA 348
358 AATCTGAATCTGCGCTCACTATGACAAATCGAATCTAGCTTAAGCGGCTTCAGTTTA 417
349 GATTAAGATTCATCTGACAGCAATTAAGATAGTAAAGTCTTGAAGATGAAAT 408
418 AGTGGAGTATTTATTTGATTTCTTCATTTGACTGCTAAACCAATGCAAGAGCTGGC 477
409 TTCACGGTCTCAACAGCCCAACAACAACGATGATTAACGCGCTTCAAGGATGAAA 468
478 ATTGACCTGATTTCCATGCGCAACAACATACCGGGAACATGGCAAGAGGATGTT 537
469 GATACCGCTTGAGAAATTTGGCAACAAACCTTGATATGTTGGAGCGGATACAGTTA 528
538 GACGGCATGGAATTTTAAGAGATGTAAGAAATGATTAATTTGGAATGGCCCAATGCA 597
529 AGTGAATCGAAAAAATTTTCTGACCAAGATGACCGGGTAAACGATGACGCTT 588
598 GCAGAGC---TGGCCAGCTTATCAATTTGTTGCAATTAATTTGACTCATTTT 654
589 GGCTTACGATGTGTCGGGAAAGTTTCGCGCTTAAAGAAATACGCGGCGTGTG 648
655 GCAGTTAGGATGATTAACCTGCGCAACAGCTAGTATGATGAACAGGATGATC 714
649 CCGGAGATCTGAAATCTTCAATCCCTATGATTTTCAAGAGCAAAAACATGCTGACAT 708
715 ACAGCAACTCCCAAGCTTTCTTAACCTGCGCAATAGTATGACCAAGATCTGATTTG 774
709 GTTGTGTGACATCACTGCGGCAAGATGATGACATATCCAAACGACCGCCAGCG 768
775 GTTATGCGCTATATCAATGCTGAAATGAGTATATTCGTAACCTGATGCAACAG 834
769 CAGCTTGCAAGAGCATGCTGATGCGGAGCTGACATCAATCGCGCCATCAACGAC 828
835 GAGGTGCTGAAAGCTTAATGATGCTGTGACAGATATTTGTTATTTGCTCCACACCCAC 894
829 GTCTTGAACCGATGATGATTAACGAAACCGCTATTTTCTACAGCTCGGCACTTT 888
895 TCCTCTGCTGCTGTGAAAATACCAAGATGGGATTTTATTTTATGCTTGGCAATTTT 954
889 GTCTTGAACCGATGATGATTAACGAAACCGCTATTTTCTACAGCTCGGCACTTT 948
955 ATTTTGAACCGATGATGATTAACGAAACCGCTATTTTCTACAGCTCGGCAATTT 1014
949 AAAAATGAAACAGGCGCTTGAAGTGAACCGATGATTAACGAAACCGCTGCA 1008
1015 AGTCAAGACAGGCTTGTATTTTACCTTAAGGCCAATGAAATGAGAGCTGATTTCCCAA 1074

OY 1009 CC 1010
DB 1075 CC 1076

RESULT 7
US-10-501-282-57

Sequence 57, Application US/10501282
Publication No. US20050203280A1

GENERAL INFORMATION:
APPLICANT: MCMICHAEL, JOHN CALHOUN

APPLICANT: ZAGURSKY, ROBERT JOHN

APPLICANT: RUSSELL, DAVID PARRISH

APPLICANT: FLETCHER, LEAH DIANE

TITLE OF INVENTION: ALLOIOCOCCUS OTITIDIS OPEN READING FRAMES (ORFS) ENCODING
FILE OF INVENTION: POLYPEPTIDE ANTIGENS, IMMUNOGENIC COMPOSITIONS AND USES THEREOF

FILE REFERENCE: AM100780 U2

CURRENT FILING DATE: 2004-07-09

PRIOR FILING DATE: 2001-11-29

PRIOR FILING DATE: 2002-11-18

PRIOR APPLICATION NUMBER: 60/426,742

PRIOR APPLICATION NUMBER: PCT/US02/36123

PRIOR FILING DATE: 2002-11-25

NUMBER OF SEQ ID NOS: 6653

SOFTWARE: PatentIn version 3.2

SEQ ID NO 57

LENGTH: 1212

TYPE: DNA

ORGANISM: Alloiococcus otitidis

FEATURE:
NAME/KEY: CDS
LOCATION: (244) ..(1209)

US-10-501-282-57

Query Match 8.5%; Score 96.4; DB 9; Length 1212;
Best Local Similarity 46.7%; Pred. No. 7, 7e-17;
Matches 393; Conservative 0; Mismatches 431; Indels 18; Gaps 2;

184 TCATTGTAGGGGATATATGATGGAGCGCTATGTTGAAAAGTAAACGAGCAAAAAGG 243
238 TCATTGTAGGGGATATATGATGGAGCGCTATGTTGAAAAGTAAACGAGCAAAAAGG 297
244 GCAGACAGTATTTTCAATATGTTGAACGATCTTGAAGCTCGGATTATGACAGA 303
298 TAGAGAGATTTTTCATATGTTAAACCAATTTGACGGTAGAGACTTGGCTGGCC 357
304 AACTTGAACCCGGTAACTATCAAAAAGATTATA-----ACAAGCA 348
358 AATCTGAATCTGCGCTCACTATGACAAATCGAATCTAGCTTAAGCGGCTTCAGTTTA 417
349 GATTAAGATTCATCTGACAGCAATTAAGATAGTAAAGTCTTGAAGATGAAAT 408
418 AGTGGAGTATTTATTTGATTTCTTCATTTGACTGCTAAACCAATGCAAGAGCTGGC 477
409 TTCACGGTCTCAACAGCCCAACAACAACGATGATTAACGCGCTTCAAGGATGAAA 468
478 ATTGACCTGATTTCCATGCGCAACAACATACCGGGAACATGGCAAGAGGATGTT 537
469 GATACCGCTTGAGAAATTTGCGAAGCAAAACCTGATATGTTGAGCGGATACAGCTTA 528
538 GACGGCATGGAATTTTAAGAGTGTAAAGTATTAATTTGAAATGGGCCATGATGCA 597
529 AGTGAATCGAAAAAATTTTCTGACCAAGATGACAGGCGTAAACGATGACGCTT 588
598 GCAGAGC---TGGCCAGCTTATCAATTTGTTGCAATTAATTTGACTCATTTT 654
589 GGCTTACGATGTGTCGGGAAAGTTTCGCGCTTAAAGAAATACGCGGCGTGTG 648
655 GCAGTTAGGATGATTAACCTGCGCAACAGCTAGTATGATGAACAGGATGATC 714
649 CCGGAGATCTGAAATCTTCAATCCCTATGATTTTCAAGAGCAAAAACATGCTGACAT 708

Db 715 ACGACCAACTCCCAAGCCTTTCTTAACCTGGCAATAGCTATGACCAAGATCTGATTGG 774
Qy 709 GTTGTGTGCACTCACTGGGCGCAAGATGATGATCCAAAGCAGCCAGCGC 768
Db 775 GTTATCGCCTATATCCAGCTGGAATAGATATATTCCTCAACCTGATGCAACCCAG 834
Qy 769 CAGCTGCAGAGCCATGCTGATGCGGAGCTGATCATCGTGGCATCATCCGAC 828
Db 835 GAGCTGGCTGAAAGCTTATGATGCTGCTGATGATGTTATTTGCTCCACACCCAC 894
Qy 829 GTCTTAGAACCGATGATATATATACGAAACCGTATTTCTACAGCCTGGCACTTT 888
Db 895 TCCCTCCCTGCTGTAAGAAATATACAGATGGGATTTATTTTATGGCTGGCAATTTT 954
Qy 889 GTCTTTGACCAAGCTGACAGAAACAGAGACAGTGCATGCTCATATCAGTGAAG 948
Db 955 ATTTTGTACCAAGGATCAAGATTCACCGATTCCTTATTTTATGACATGATATCAAT 1014
Qy 949 AAAATGGAACAGCGCGCTTGAAGTGAACCGATGATATCCATGAGCGACCTGCA 1008
Db 1015 AGTCCAGACCAAGTTCCTTTTACCTTAAGCCCAATGAAATGAGCTGATTTCCCAA 1074
Qy 1009 CC 1010
Db 1075 CC 1076

RESULT 8

US-10-282-59
; Sequence 59, Application US/10501282
; Publication No. US20050203280A1
; GENERAL INFORMATION:
; APPLICANT: MCMICHAEL, JOHN CALHOUN
; APPLICANT: ZAGORSKY, ROBERT JOHN
; APPLICANT: RUSSELL, DAVID PARRISH
; APPLICANT: FLETCHER, LEAH DIANE
; TITLE OF INVENTION: ALLOIOCCUS OTITIDIS
; TITLE OF INVENTION: POLYPEPTIDE ANTIGENS, IMMUNOGENIC COMPOSITIONS AND USES THEREOF
; FILE REFERENCE: AM100780 L2
; CURRENT APPLICATION NUMBER: US/10/501,282
; PRIOR FILING DATE: 2004-07-09
; PRIOR APPLICATION NUMBER: 60/333,777
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 60/426,742
; PRIOR FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: PCT/US02/36123
; PRIOR FILING DATE: 2002-11-25
; NUMBER OF SEQ ID NOS: 6653
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 59
; LENGTH: 1212
; TYPE: DNA
; ORGANISM: Alloiococcus otitidis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (16)..(1209)
US-10-501-282-59

Query Match 8.5%; Score 96.4; DB 9; Length 1212;
Best Local Similarity 46.7%; Pred. No. 7,7e-17;
Matches 393; Conservative 0; Mismatches 431; Indels 18; Gaps 2;

Qy 184 TCATTGTGAGCGATATATGATGAGCGCTATGTTGAAAAAGTAAAGGAGAAAAAGGG 243
Db 238 TCATTGTGAGCGATATATGATGAGCGCTATGTTGAAAAAGTAAAGGAGAAAAAGGG 297
Qy 244 GCAGACGATATTTTCAATATGTAACCGATCTTAAAGCTGCGATATGATGACAGCA 303
Db 298 TACGAGGAATTTTACATATGTAACCGATCTTAAAGCTGCGATATGATGACAGCA 357
Qy 304 AACTTGAAGAACCGTAACTATCAAAAAGATTTAA-----ACAAGCA 348

Db 358 AATCTGATATCTGCGCTACCTATGACATGAACTTACCTTAAGCCGCTTCAAGTTTA 417
Qy 349 GATAAGAGATTCATCTGACAGCAATATGAAATCAGTGAAGTCTTGAAGATATGAAT 408
Db 418 AGTGAGGATTTTATTTGATTTCTTCAATGATCTGATCAAGGATGCAAGAGCTGGC 477
Qy 409 TTCAAGGTTCTCAACAGCGCCAAACACACGCAATGATTAAGCGCTTCAAGGATGAA 468
Db 478 ATTACCTAGTTTCATGCGCAACCAACATACCGGCAATGCGCAAGAGGATGATGTT 537
Qy 469 GATACGCTTGAAGATTTGGAAGCAAAACCTGATATCGTTGAGAGGGATACAGTTA 528
Db 538 GACGCAATGAAATTTTAAAGATGTAAGATTAATATGAAATGAGGCAATGATGCA 597
Qy 529 AGTATGCGAAAAAGAAATTTTGTATACAGAAAGTCAACGAGGAGTATGCAACGCTT 588
Db 598 GCAGAGC---TGCCAGAGCTTATCAATTTGTGCGAATATTTGACTTACTCATTTT 654
Qy 589 GCGTTTACGATGTGTCGGGAAAGTTTCCGCTTAAAGAAATACGCGGCGTGTG 648
Db 655 GCACTTGCATGATGATTAACCTGCGCAAGCAGCTATGATGATGAAACGAGTGTACTC 714
Qy 649 CCGCAGATCTGAAATCTTCAATCCCTATGATTTTCAAGAGCAAAAAACATGTCATAT 708
Db 715 ACGACCAACTCCCAAGCCTTTCTTAACCTGCGCAATGATGACCAAGATCTGATTTG 774
Qy 709 GTTGTGTGCACTCACTGGGCGCAAGATGATGACATGATCCAAAGCAGCCAGCGC 768
Db 775 GTTATCGCCTATATCCATGCTGGAATAGATATTTGCTCAACCTGATGCAACCAAC 834
Qy 769 CAGCTGCAGAGCCATGCTGATGCGGAGCTGACATCATCGTGGCATCATCGCAC 828
Db 835 GAGCTGGTGAAGACTTAATGATGCTGTCGATGATGATGATTTATTTCTCCCAACCCAC 894
Qy 829 GTCTTAGAACCGATGATATATATACGAAACCTCATTTTCTACAGCTGCGCACTTT 888
Db 895 TCCCTCCCTGCTGTAAGAAATATCAAGATGGAATTTTATTTTATGCTTGGGCAATTTT 954
Qy 889 GTCTTAGAACAGGCTGAGAGAAACAGACAGTGCATCGTTCATATCATCCTGAAG 948
Db 955 ATTTTGCAGAGCGATGCAAGTTCACCGATTCCTTATTTTATGACATGATATCAAT 1014
Qy 949 AAAATGGAACAGCGCGCTTGAAGTGAACCGATGATATCCATGAGCGACCTGCA 1008
Db 1015 AGTCCAGACCAAGTTCCTTTTACCTTAAGCCCAATGAAATGAGCTGATTTCCCAA 1074
Qy 1009 CC 1010
Db 1075 CC 1076

RESULT 9

US-10-501-282-6651/C
; Sequence 6651, Application US/10501282
; Publication No. US20050203280A1
; GENERAL INFORMATION:
; APPLICANT: MCMICHAEL, JOHN CALHOUN
; APPLICANT: ZAGORSKY, ROBERT JOHN
; APPLICANT: RUSSELL, DAVID PARRISH
; APPLICANT: FLETCHER, LEAH DIANE
; TITLE OF INVENTION: ALLOIOCCUS OTITIDIS
; TITLE OF INVENTION: POLYPEPTIDE ANTIGENS, IMMUNOGENIC COMPOSITIONS AND USES THEREOF
; FILE REFERENCE: AM100780 L2
; CURRENT APPLICATION NUMBER: US/10/501,282
; PRIOR FILING DATE: 2004-07-09
; PRIOR APPLICATION NUMBER: 60/333,777
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 60/426,742
; PRIOR FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: PCT/US02/36123
; PRIOR FILING DATE: 2002-11-25
; NUMBER OF SEQ ID NOS: 6653
; SOFTWARE: PatentIn version 3.2

SEQ ID NO 6651
 LENGTH: 1754382
 TYPE: DNA
 ORGANISM: *Alliobococcus oolitidis*
 US-10-501-282-6651

Query Match 8.5%; Score 96.4; DB 9; Length 1754382;
 Best Local Similarity 46.7%; Pred. No. 3.9e-15;
 Matches 393; Conservative 0; Mismatches 431; Indels 18; Gaps 2;

184 TCATTGTAGGCGAATATTATGATGGAGCGCTATGTTGAAAAAGTAACGAGCAAAAAGG 243
 24408 TCATTGTGGGCGAATATCCTTGAGTGGCTATATGGGGAATGGGGAACGATG 24349
 244 GCAGACATATTATTCATATGTTAACCGCATCTTTAAGCCTCGAATTATGTACAG 303
 24348 TACGAGAAATTTTACATATGTTAACCAATTTGACGGTAGAGACTGTGCTGCC 24289
 304 AACTTTGAAAAACCGGTAACTATCAAAAGATTATA-----ACNAGCA 348
 24288 AATCTTGATCTGCCGTCACTATGACATCGAATCTGTTAGCCGGCTTCAAGTTTA 24229
 349 GATAAGAGATTCATCTGAGACGAAATAGAAATGAGTAAGTAAGTCTTGAAGGATGA 408
 24228 AGTGAAGTATTATTATGATTTCTTCATTGACTGCTCAAGCCATGCAAGAGCTGGC 24169
 409 TTCAGGTTCTCACAAGCGCCAAACCAACGCAATGATTAAGCGCTTCAGGGCATGAA 468
 24168 ATTGACCTAGTTTCATGCGCCAAACCAATACCGGACATGGGCAAGAGGTATGTT 24109
 469 GATACGCTTGAGAAATTTGCGAAGCAAACTTGATATGTTGAGCGGGAATACAGCTTA 528
 24108 GACGCGATGAAATTTTAAAGATGTAGAAATGAAATTAATATTATGATGGCCATGATCA 24049
 529 AGTATGCGAAAAAAGAAATTTCTACGAAAGTCAACGGGGTAAAGTCAAGTCAAGCTT 588
 24048 GCAGAGC---TGCCAAAGCTTATCAATTTGTTGCGAATTAATTTGACTCTCATTTT 23992
 589 GCGTTTACCGATGTGTCGCGGAAAGGTTTCGCGGCTTAAAGAAATGACGCGGCGGTCTG 648
 23991 GCAAGTTAGGATGTGATTAACCTGCGCAAGCAGCTAGTAGATGAACCAAGGTACTC 23932
 649 CCGCGAGATCTGAAATCTTCACTATGATTTTCAAAAGCGAAAAAATCATGCTGACAT 708
 23931 ACGACCAACTCCCAAGCTTTCTTAACCTGCGCAATAGCTATGACCAAGATCTGATTTG 23872
 709 GTTGTGTGACGACACACTGGGCGCAAGTATGACATGATCCAAAGCAGCGCGC 768
 23871 GTTATGCGCTATATCATGCTGGAATGAGTATATTCGTCACCTGATGCAACCAACG 23812
 769 CAGCTTGCAAGACCATGCTGAGCGGAGCTGACATCATCGTCCGCAATCCGCAC 828
 23811 GAGCTGCGTGAAGCTTAATGATGCTGAGTGAATGTTGTTGCTCCCAACCCAC 23752
 829 GTCTTAGAACCGATTGAAGTATTAACGAAACCGTCAATTTTCAAGCCTCGCAACTTT 888
 23751 TCCCTCCTGCTGTGTAATAAATACCAAGATGGAATATTTTATGAGCTTGCGCAATTT 23692
 889 GTCTTGACCAAGCTGAGAGAGAAACAGAGACAGTCACTGTTCAAGTACCTGAAAG 948
 23691 AATTTTGAACCAAGGATGCAAGATTCACCGAATTCGTTAATTTTGAAGATGATATCAAT 23632
 949 AAAAAAGGAACAGCGCGCTTGAAGTACACCGATGATTCATGCAAGGACACCTGCA 1008
 23631 AGTCCAGACCAAGGTTGTTTACCTTAGGCAATGAATAAGAGCTGTATTTCCAAA 23572
 1009 CC 1010
 23571 CC 23570

RESULT 10
 US-10-398-221-10/c

Sequence 10, Application US/10398221
 Publication No. US20040018514A1
 GENERAL INFORMATION:
 APPLICANT: KUNST, Frederik
 APPLICANT: GLASER, Philippe
 TITLE OF INVENTION: *Listeria innocua*, genome and applications
 FILE REFERENCE: 344 702 - US
 CURRENT APPLICATION NUMBER: US/10/398,221
 CURRENT FILING DATE: 2003-03-27
 PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
 PRIOR FILING DATE: 2001-10-04
 PRIOR APPLICATION NUMBER: FR 00/12 697
 PRIOR FILING DATE: 2000-10-04
 NUMBER OF SEQ ID NOS: 4025
 SOFTWARE: PatentIn version 3.0
 SEQ ID NO 10
 LENGTH: 1163020
 TYPE: DNA
 ORGANISM: *Listeria innocua*
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (1)..(end)
 OTHER INFORMATION: n can be any nucleotide: a, g, c or t/u
 US-10-398-221-10

Query Match 6.8%; Score 77.8; DB 7; Length 1163020;
 Best Local Similarity 57.0%; Pred. No. 8.2e-10;
 Matches 142; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

649 CCGCGAGATCTGAAATCTTCACTATGATTTGATTCAGAGCGAAAAAATCATGTCACAT 708
 446758 CCAGCTATCTTGAAGAAATGTAAGACGTTAAAAAATTAAGAAAGATAGCTC 446699
 709 GTTGTGTGACGTCACACTGGGCGCAAGATGACAAATGATCCAAAGACCGCCAGCG 768
 446698 GTTATTTGTAATATGACACTGGGCGCTCGAATATCGTGAACCTCCAAACCAACT 446639
 769 CAGCTTGCAAGGCGCATGCTGATCGGAGCGTGAATCATGCTGCGCATCATCCGCAC 828
 446638 CAATTGACATGCTAATTTTAAAGATGCTGCTGATATATTTATGAGACTATCCAC 446579
 829 GTCTTAGAACCGATTGAAGTATTAACGAAACCGTCAATTTTCAAGCCTCGCAACTTT 888
 446578 CCGCTGGAAGATGCAAAATTAATAAGATTAATATTTGATATACGATGGCGATTTT 446519
 889 GTCTTTGAC 897
 446518 GCTTTTGGC 446510

RESULT 11
 US-10-398-221-2058/c
 Sequence 2058, Application US/10398221
 Publication No. US20040018514A1
 GENERAL INFORMATION:
 APPLICANT: KUNST, Frederik
 APPLICANT: GLASER, Philippe
 TITLE OF INVENTION: *Listeria innocua*, genome and applications
 FILE REFERENCE: 344 702 - US
 CURRENT APPLICATION NUMBER: US/10/398,221
 CURRENT FILING DATE: 2003-03-27
 PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
 PRIOR FILING DATE: 2001-10-04
 PRIOR APPLICATION NUMBER: FR 00/12 697
 PRIOR FILING DATE: 2000-10-04
 NUMBER OF SEQ ID NOS: 4025
 SOFTWARE: PatentIn version 3.0
 SEQ ID NO 2058
 LENGTH: 3011208
 TYPE: DNA
 ORGANISM: *Listeria innocua*
 US-10-398-221-2058

Query Match 6.8%; Score 77.8; DB 7; Length 3011208;
Best Local Similarity 57.0%; Pred. No. 1.4e-09;
Matches 142; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

QY 649 CCGGAGATCTGAAATCTTCAATCCATATTCAGATTCAGAGGAGGAAAAATCATGCTGACATTT 708
DB 537024 CCAGCCATCTTTGAAAGAAATGTAAAGACGTTAAAAATATTAAGAAAGAAATACGCTC 536965
QY 709 GTTGTGGGAGCTGACATCTGGGGCCAAAGATATGCAATGATCCAAAGACCGGAGCGC 768
DB 536964 GTTATTTTATATGACCTGGGGCGTCAATATGCTGTAACCTCCAACTCAACTCAACTCAACT 536905
QY 769 CAGCTTGCAAGAGCCATCTGATGCGGAGCTGACATCATCTGCGCCATCATCCGAC 828
DB 536904 CAATTGGACATGCTATTTTATGATCTGCTGATATTAATTAAGAGAGCTCATCCAC 536845
QY 829 GTCTTAGAACGATTAAGATATTAACGAAACCGTCAATTTTCTACAGCCTCGCAACTTT 888
DB 536844 CGCCTGAAAGATATGAAATATTAAGATTAATATTTGATACAGATGCGGCAATTTT 536785
QY 889 GTCTTTGAC 897
DB 536784 GCTTTTGGC 536776

RESULT 12

US-10-398-221-1526
; Sequence 1526, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; PRIOR FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1526
; LENGTH: 885
; TYPE: DNA
; ORGANISM: Listeria monocytogenes-4B
US-10-398-221-1526

Query Match 6.5%; Score 74; DB 7; Length 885;
Best Local Similarity 54.4%; Pred. No. 2.2e-10;
Matches 149; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

QY 622 GCTAAAAAGAAATACGCGGGCGTCTGCGCCGAGATCTGAAATCTTCAATCCATATGATT 681
DB 40 GCAGACAGCATACACGCGCATATCCATGCTGATTAATGAAGAAATGTACGACGCTG 99
QY 662 TCAGAGGAGAAAAAATGCTGACATTTGTTGTGTCAGTCACTGCGGCGCAAGATAT 741
DB 100 AAGACATACAAAAAAGAAACTCTGCTGATATGATTAATACCATGCGGCGCTCAATAT 159
QY 742 GACATATGATCCAAAGACCGGAGCGGAGCTTGGCAAGCCATGCTGATGCGGAGCT 801
DB 160 CGCGAAACACCAAGGATTTATCAACCAATTTGCTCATGCAATTTTATGATGCTGCT 219
QY 802 GACATCATCTGCGGAGCATATCCGACGCTTAAAGACGATTAAGATTAATTAACGAAAC 861
DB 220 GATATTAATTAATGCTCTACACCTCATGCGCTAGAAAGGCTTGAAGAAATATTAAGATATA 279
QY 862 GTCAATTTTCTACAGCTCGGCACTTTGCTTTG 895
DB 280 TACATCGTTTATAGCATGGCGCACTTTGCTTTG 313

RESULT 13
US-10-398-221-3371
; Sequence 3371, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; PRIOR FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3371
; LENGTH: 2282
; TYPE: DNA
; ORGANISM: Listeria monocytogenes 4b
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(end)
; OTHER INFORMATION: n can be any nucleotide: a,g,c or t/u
US-10-398-221-3371

Query Match 6.5%; Score 74; DB 7; Length 2282;
Best Local Similarity 54.4%; Pred. No. 3.6e-10;
Matches 149; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

QY 622 GCTAAAAAGAAATACGCGGGCGTCTGCGCCGAGATCTGAAATCTTCAATCCATATGATT 681
DB 40 GCAGACAGCATACACGCGCATATCCATGCTGATTAATGAAGAAATGTACGACGCTG 99
QY 682 TCAGAGGAGAAAAAATGCTGACATTTGTTGTGTCAGTCACTGCGGCGCAAGATAT 741
DB 100 AAGACATACAAAAAAGAAACTCTGCTGATATGATTAATACCATGCGGCGCTCAATAT 159
QY 742 GACATATGATCCAAAGACCGGAGCGGAGCTTGGCAAGCCATGCTGATGCGGAGCT 801
DB 160 CGCGAAACACCAAGGATTTATCAACCAATTTGCTATGATGATTAAGATGCTGCTGCT 219
QY 802 GACATCATCTGCGGAGCATATCCGACGCTTAAAGACGATTAAGATTAATTAACGAAAC 861
DB 220 GATTAATTAATGCTCTACACCTCATGCGCTAGAAAGGCTTGAAGAAATATTAAGATATA 279
QY 862 GTCAATTTTCTACAGCTCGGCACTTTGCTTTG 895
DB 280 TACATCGTTTATAGCATGGCGCACTTTGCTTTG 313

RESULT 14

US-10-470-048B-514
; Sequence 514, Application US/10470048B
; Publication No. US20050037444A1
; GENERAL INFORMATION:
; APPLICANT: MEINKE ET AL.
; TITLE OF INVENTION: A METHOD FOR IDENTIFICATION, ISOLATION AND PRODUCTION OF
; FILE REFERENCE: SONN:03505
; CURRENT APPLICATION NUMBER: US/10/470,048B
; PRIOR FILING DATE: 2003-07-25
; NUMBER OF SEQ ID NOS: 603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 514
; LENGTH: 1071
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-10-470-048B-514

Query Match 4.8%; Score 54.2; DB 8; Length 1071;
Best Local Similarity 47.4%; Pred. No. 0.00014;

Matches 229; Conservative 0; Mismatches 248; Indels 6; Gaps 2;

QY 465 GAAAGATACGCTTGAGAAATTTGCGAAGCAAACTTGATATCGTTGGACGGGATACAG 524
 DB 432 GACAGCAATGATGAGAAAAATAGATTTCCAGACAGATTATTTTAAACAGTAAATG 491
 QY 535 CTTAAGTATGCGAAAAAATTTTCATACAGAAAGTCAACGGGGTACAGATTGCAAC 584
 DB 492 TTCAATCCAAATTAATAGTAAACGTACACAGACATTAAGTAAAGTAAATAGCTAA 551
 QY 585 GCTTGCGCTTACCGATGTGTCCGGGAAAGTTTCGGCGCTAAAGAAATACCGCGGCGT 644
 DB 552 TGTTCATTTACCGATATGAAATCTACATTAATACTTTAAAAAACAAGACG---TC 608
 QY 645 GCTGCCCGCAGATCCTGAAATCTTCATCCCTATGATTTGAGAAGCGAAAAACATGCTGA 704
 DB 609 AATTAGTTAGATCCAGCTATATTTATCTTTAATTAATAAATAATTAAGAAAAATATGA 668
 QY 705 CATTTGTTGTGCGACATCACTGGGGCCAGAGATGACATGATCCAAACGACCGCA 764
 DB 669 TTACGTCGATGCAATGTATGATTTGGGGATACCTAATGAACGAATGTGACTACATCA 728
 QY 765 GGGCCAGCTTGCAAGACCATGTCTGATGCGGAGCTGACATCATCGTGGCCATCATCC 824
 DB 729 AAAAGATATGACATGCGTTACCGAATCTGTGCGAGATGTCATTAATGTCATTAATAC 788
 QY 825 GCACGCTTTAGAACCGATTTGAAGTATATACGGAACCGTCAATTTTCTACAGCCTCGGCA 884
 DB 789 ---AGTTATTCAAAAAGTTGAAAATATATAAGCAAGCCATATTTTATATAGTTAGTAA 845
 QY 885 CTTTGTCTTTGACCAAGCTGCGACGAGAACAGACAGTGCACCTGTCATATACCT 944
 DB 846 CACACGCTGTATTAATCTTTATCAAAAATCAGAAAGATGATGTACACAGACTG 905
 QY 945 GAA 947
 DB 906 GAA 908

RESULT 15
 US-10-724-972A-1055
 ; Sequence 1055, Application US/10724972A
 ; Publication No. US20040147734A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Doucette-Stamm, Lynn
 ; APPLICANT: Bush, David
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: PATH03-16
 ; CURRENT APPLICATION NUMBER: US/10/724,972A
 ; CURRENT FILING DATE: 2003-12-01
 ; PRIOR APPLICATION NUMBER: 09/450,969
 ; PRIOR FILING DATE: 1999-11-29
 ; PRIOR APPLICATION NUMBER: 09/134,001
 ; PRIOR FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 7544
 ; SEQ ID NO 1055
 ; LENGTH: 1101
 ; TYPE: DNA
 ; ORGANISM: S. epidermidis
 US-10-724-972A-1055

Query Match 4.88; Score 54.2; DB 7; Length 1101;
 Best Local Similarity 47.48; Pred. No. 0.00014;
 Matches 229; Conservative 0; Mismatches 248; Indels 6; Gaps 2;

QY 465 GAAAGATACGCTTGAGAAATTTGCGAAGCAAACTTGATATCGTTGGACGGGATACAG 524
 DB 459 GACAGCAATGATGAGAAAAATAGATTTCCAGACAGATTATTTTAAACAGTAAATG 518

QY 525 CTTAAGTATGCGAAAAAATTTCTACACAGAAAGTCAACGGGGTACAGATTGCAAC 584
 DB 519 TTCAATCCAAATTAATAGTAAACGTACACAGACATTAAGSTAAATAATAGCTAA 578
 QY 585 GCTTGCGCTTACCGATGTGTCCGGGAAAGTTTCGGCGCTAAAGAAATAGCGCGGCGT 644
 DB 579 TGTTCATTTACCGATATGCAATCTAATTAATACTTTTAAAAAACAAGACG---TC 635
 QY 645 GCTGCCCGCAGATCCTGAAATCTTCATCCCTATGATTTTCAAGAGGAAAAACATGCTGA 704
 DB 636 AATTAGTTAGATCCAGCTATATTTATCTTTAATTAATAAATAATTAAGAAAAATATGA 695
 QY 705 CATTTGTTGTGCGACATCACTGGGGCCAGAGATGACATGATCCAAACGACCGCA 764
 DB 696 TTACGTCGATGCAATGTATGATTTGGGGATACCTAATGAAGAAATGTGACTACACGCTCA 755
 QY 765 GGGCCAGCTTGCAAGACCATGTCTGATGCGGAGCTGACATCATCGTGGCCATCATCC 824
 DB 756 AAAAGATATGACATGCGTTAGCGAATGCTGTGCGAGATGTCATTAATGTCATTAATAC 815
 QY 825 GCACGCTTTAGAACCGATTTGAAGTATATACGGAACCGTCAATTTTCTACAGCCTCGGCA 884
 DB 816 ---AGTTATTCAAAAAGTTGAAAATATATAAGCAAGCCATATTTTATAGTTAGTAA 872
 QY 885 CTTTGTCTTTGACCAAGCTGCGACGAGAACAGACAGTGCACCTGTCATATACCT 944
 DB 873 CACACGCTGTATTAATCTTTATCAAAAATCAGAAAGATGATGTACACAGACTG 932
 QY 945 GAA 947
 DB 933 GAA 935

Search completed: February 27, 2006, 07:58:43
 Job time : 1259.33 secs

THIS PAGE BLANK (USPTO)


```
OY 181 GCCTATTGTAGGCGATATTATGATGAGGACGCTATGTTGAAAAAGTAACGAGCAAAA 240
Db 181 GCCTATTGTAGGCGATATTATGATGAGGACGCTATGTTGAAAAAGTAACGAGCAAAA 240
OY 241 GGGGCGACAGATATTTTCAATATGTTGACCGATCTTTAGAGCTCGGATTTATGACA 300
Db 241 GGGGCGACAGATATTTTCAATATGTTGACCGATCTTTAGAGCTCGGATTTATGACA 300
OY 301 GGAACCTTTGAAAAACCCGGTAACCTATCAAAAAGATTATAACAAGCAGATTAAGAGTT 360
Db 301 GGAACCTTTGAAAAACCCGGTAACCTATCAAAAAGATTATAACAAGCAGATTAAGAGTT 360
OY 361 CATGCGACAGAGATTAAGATTAAGTAAAGCTTTGAAAGATATGATTAATTTACGGTTCTC 420
Db 361 CATGCGACAGAGATTAAGATTAAGTAAAGCTTTGAAAGATATGATTAATTTACGGTTCTC 420
OY 421 AACAGCGCCAAACAACGACGATGATTAACGCGTTCAAGGCGATGAAAGATACGTTGGA 480
Db 421 AACAGCGCCAAACAACGACGATGATTAACGCGTTCAAGGCGATGAAAGATACGTTGGA 480
OY 481 GAATTTGGGAAAGCAAAACCTTGATATCGTTGAGCGGGATACAGCTTAAGTATGCGGAA 540
Db 481 GAATTTGGGAAAGCAAAACCTTGATATCGTTGAGCGGGATACAGCTTAAGTATGCGGAA 540
OY 541 AAGAAAAATTTGTTACAGAAAGTCAACGCGGTAACGATTTGCAACGCTTGGCTTTACGAT 600
Db 541 AAGAAAAATTTGTTACAGAAAGTCAACGCGGTAACGATTTGCAACGCTTGGCTTTACGAT 600
OY 601 GTGTCCGGGAAAAGTTTCCGCGCTAAAAAGATACGCCGGGCGTCTGCCGCAAGATCTT 660
Db 601 GTGTCCGGGAAAAGTTTCCGCGCTAAAAAGATACGCCGGGCGTCTGCCGCAAGATCTT 660
OY 661 GAAATCTTCATCCCTATGATTTCAAGACGCAAAAAACATGCTGATGATTTGTTGTGAG 720
Db 661 GAAATCTTCATCCCTATGATTTCAAGACGCAAAAAACATGCTGATGATTTGTTGTGAG 720
OY 721 TCACACTGGGGCCAGAGTATGACAAATGATCCAAAGCAGCCGACGCGCAAGCTTGCAAGA 780
Db 721 TCACACTGGGGCCAGAGTATGACAAATGATCCAAAGCAGCCGACGCGCAAGCTTGCAAGA 780
OY 781 GCCATGTCTGATGCGGAGCTGACATCATCGTGGCCATCATCGCACGCTTTAGAACCG 840
Db 781 GCCATGTCTGATGCGGAGCTGACATCATCGTGGCCATCATCGCACGCTTTAGAACCG 840
OY 841 ATTGAAGTATATACGGAACCGCTATTTTCTACAGCTCGGCAACTTTGTCTTTGACCA 900
Db 841 ATTGAAGTATATACGGAACCGCTATTTTCTACAGCTCGGCAACTTTGTCTTTGACCA 900
OY 901 GGGTGGACGAGAAACAAGACAGTGCACGTTTCAATGATCACTGAAGAAAAATGAAACA 960
Db 901 GGGTGGACGAGAAACAAGACAGTGCACGTTTCAATGATCACTGAAGAAAAATGAAACA 960
OY 961 GGCCGCTTTGAAGTACACCGATTCATTCATGAAAGGACACCTGCACTGTGAAAAAA 1020
Db 961 GGCCGCTTTGAAGTACACCGATTCATTCATGAAAGGACACCTGCACTGTGAAAAAA 1020
OY 1021 GACAGCCTTTAAACAGAAAAACCTTATTTGGCGAAGTGAAGAAAGACTTAAATTTGCGTTGG 1080
Db 1021 GACAGCCTTTAAACAGAAAAACCTTATTTGGCGAAGTGAAGAAAGACTTAAATTTGCGTTGG 1080
OY 1081 AAAGTAGAAGACGGAAGAACTGAGTTGATTAATGATCAATAGACAAACTAAACTTAA 1140
Db 1081 AAAGTAGAAGACGGAAGAACTGAGTTGATTAATGATCAATAGACAAACTAAACTTAA 1140
```

RESULT 2

```
US-10-793-626-4412
; Sequence 4412, Application US/10793626
; Publication No. US2005025478A1
; GENERAL INFORMATION:
; APPLICANT: KIMBERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
```

```
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4412
; LENGTH: 3523
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-10-793-626-4412
```

```
Query Match 4.2%; Score 47.6; DB 8; Length 3523;
Best Local Similarity 54.6%; Pred. No. 0.016;
Matches 95; Conservative 0; Mismatches 79; Indels 0; Gaps 0;
```

```
OY 686 AAGGAAAAACATGCTGACATGTTGTTGTCAGTCACTGGGCGCAAGATATGACA 745
Db 2635 AAGGAAAAACATGCTGACATGTTGTTGTCAGTCACTGGGCGCAAGATATGACA 2634
OY 746 ATGATCCAAACGACCGCCAGCCAGCTTGCAAGACCATGCTGATCGGAGCTGACA 805
Db 2695 ATGATCCAAACGACCGCCAGCCAGCTTGCAAGACCATGCTGATCGGAGCTGACA 804
OY 806 TCATGCTGGCCATCATTCGCGACGCTTTAGAACCGATTTGAAGTATTAACGGA 859
Db 2755 TAGTTTAAAGTACATCTCATGATGATTCACCAAGTAAATGGGTTGCGTA 2808
```

RESULT 3

```
US-11-114-798-55/C
; Sequence 55, Application US/11114798
; Publication No. US20060035246A1
; GENERAL INFORMATION:
; APPLICANT: WU, RINA
; APPLICANT: MARQUEZ, ABBEY
; TITLE OF INVENTION: CHROMOGENIC IN SITU HYBRIDIZATION METHODS, KITS, AND
; FILE REFERENCE: 0618.011.0004
; CURRENT APPLICATION NUMBER: US/11/114,798
; CURRENT FILING DATE: 2005-04-26
; PRIOR APPLICATION NUMBER: 10/173,525
; PRIOR FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: 09/952,851
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 55
; LENGTH: 173120
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-11-114-798-55
```

```
Query Match 3.7%; Score 42.4; DB 9; Length 173120;
Best Local Similarity 50.5%; Pred. No. 2.4;
Matches 103; Conservative 0; Mismatches 101; Indels 0; Gaps 0;
```

```
OY 368 AGACGATTAAGAAATCAGTGAAGCTTTGAAGCATATGAATTTACAGGTTTCAACGCG 427
Db 156108 AAGGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 156049
OY 428 CCAACAACGACGATGATTAACGCGCTTCAGGCGCATGAAGATACGCTTGAGAGATTG 487
Db 156048 AAGGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 155989
OY 488 CGAAGCAAAACCTGATATCGTTGAGCGGATACAGCTTAAGTATGCGAAAAAGAAA 547
```


Db 1744 TCTGTGCTCTGGGGAGATCTTGAAAAACAGCCTAACCTCCCTGATGTCAAGTGGC 1803
Qy 363 TCTGCAGACGAATAGAAATCAAGAAAGTCTTGAAGATATGAATTTACAGGTTCTCAA 422
Db 1804 TGCAGTGTGAAGATTACTTGAAGATTTCTTGACGCAATGAATGATTCAGACCCAG 1863
Qy 423 CAGCGCCCAACCAACCAATGAT 447
Db 1864 TGCCACGAGGAATAGCAATTGAT 1888

RESULT 7

US-09-925-065A-700467/c
; Sequence 700467, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 700467
; LENGTH: 620
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-700467

Query Match 3.4%; Score 39; DB 6; Length 620;
Best Local Similarity 49.3%; Pred. No. 1.6;
Matches 102; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

Qy 225 AGTACGAGCAAAAAGGGGACAGATATTTTCAATATGTGAACCGATTTTAAGC 284
Db 370 ATTAATGTCCCACTTAAGTGGGCAATGTAATGATGGGATACATGACCCCTTGTCTC 311
Qy 285 CTCGATTATGTAAGAGAAACCTTGAAAACCGGTAACCTATCAAAAGAAATTATAACA 344
Db 310 AGGGATGTTTCCACAAACATCTTGSATBACCCTGTCCCGATCTGTTGGTTCTAAC 251
Qy 345 AGCAGATAAGAGATTCACTGTCAGACAGAAATAGAAATCAGTGAAGCTTGAAGATAT 404
Db 250 CATAAATGATGGGTTGAGCATGGAGGATATAGATAGATTAAGAAACAGAGATGT 191
Qy 405 GAATTTACGGTTCTCAACAGCGCCAA 431
Db 190 GTACAACCTCGGGGCACATCATGGCCAA 164

RESULT 8

US-09-925-065A-700468/c
; Sequence 700468, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08

; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 700468
; LENGTH: 620
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-700468

Query Match 3.4%; Score 39; DB 6; Length 620;
Best Local Similarity 49.3%; Pred. No. 1.6;
Matches 102; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

Qy 225 AGTACGAGCAAAAAGGGGACAGATATTTTCAATATGTGAACCGATTTTAAGC 284
Db 370 ATTAATGTCCCACTTAAGTGGGCAATGTAATGATGGGATACATGACCCCTTGTCTC 311
Qy 285 CTCGATTATGTAAGAGAAACCTTGAAAACCGGTAACCTATCAAAAGAAATTATAACA 344
Db 310 AGGGATGTTTCCACAAACATCTTGSATBACCCTGTCCCGATCTGTTGGTTCTAAC 251
Qy 345 AGCAGATAAGAGATTCACTGTCAGACAGAAATAGAAATCAGTGAAGCTTGAAGATAT 404
Db 250 CATAAATGATGGGTTGAGCATGGAGGATATAGATAGATTAAGAAACAGAGATGT 191
Qy 405 GAATTTACGGTTCTCAACAGCGCCAA 431
Db 190 GTACAACCTCGGGGCACATCATGGCCAA 164

RESULT 9

US-09-925-065A-55305/c
; Sequence 55305, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 55305
; LENGTH: 549
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-55305

Query Match 3.3%; Score 38; DB 6; Length 549;
Best Local Similarity 59.1%; Pred. No. 2.9;
Matches 65; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Qy 1 ATGAAAAAGAACTGAGCTTTCATGAAAGCTGTAAAGCTGACAAAAACGCAAAAAAAG 60

Db 187 ATGAATCCGACTGAGCTGTATGAGCCTTACTTAACGTGAGTTACAGTATATTTA 128
OY 61 AAACCAATAGACAGTATTTATTCGATTCGATCGTTTGTCTTAT 110
Db 127 AAAACAGAGAAAGATTTTCTTTCTTTCTTTCTTTTCTTTT 78

RESULT 10
US-10-467-5053
; Sequence 5053, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqMan99, version 1.04
; SEQ ID NO 5053
; LENGTH: 771
; TYPE: DNA
; ORGANISM: Neisseria gonorrhoeae
US-10-467-5053

Query Match 3.3%; Score 37.6; DB 8; Length 771;
Best Local Similarity 56.5%; Pred. No. 4.3;
Matches 70; Conservative 0; Mismatches 54; Indels 0; Gaps 0;
OY 768 CCAGCTTGCAAGACCATGTGATGCGGAGCTGACATCGTCGCGCATTCGCA 827
Db 426 CGAGTTTGGCGGCGTCCGCGTGCATTCGCGCGGATGTCGATTCGCGCGAGCGCGCA 485
OY 828 CGCTTGAACCGATTTGAGTATATACGGAACCGCATTTTCTTAAGCTCTGGCACTT 887
Db 486 CGTTACTCGCGCGCTGCACTTACACGACCGCTTCACTCTTACAGCGCGCACTT 545
OY 888 TGTC 891
Db 546 TGCC 549

RESULT 11
US-10-467-5055/C
; Sequence 5055, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqMan99, version 1.04
; SEQ ID NO 5055
; LENGTH: 1065
; TYPE: DNA
; ORGANISM: Neisseria gonorrhoeae
US-10-467-5055

Query Match 3.3%; Score 37.6; DB 8; Length 1065;
Best Local Similarity 56.5%; Pred. No. 5;
Matches 70; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

OY 768 CCAGCTTGCAAGACCATGTGATGCGGAGCTGACATCATCTCGGCATATCCGCA 827
Db 448 CGAGTTTGGCGGCTCCCGCTCGATTCGCGCGGAGTGTATTCGCGCGAGCGGCA 389
OY 828 CGCTTGAACCGATTTGAGTATATACGGAACCGCATTTTCTTAAGCTCTGGCACTT 887
Db 388 CGTTACTCGCGCGCTGCACTTACACGACCGCTTCACTCTTACAGCGCGCACTT 329
OY 888 TGTC 891
Db 328 TGCC 325

RESULT 12
US-09-925-065A-144642/C
; Sequence 144642, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 144642
; LENGTH: 543
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-144642

Query Match 3.2%; Score 36.6; DB 6; Length 543;
Best Local Similarity 52.3%; Pred. No. 6.9;
Matches 81; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

OY 412 ACGTTTCTCAACAGCGCAACCAACGATGATACGCGCTTACGCGCATGAAAGT 471
Db 301 AAGATTTCAGAAATAATTAATAACAGATTAGAGTACATATACAGGACCTGTAAT 242
OY 472 ACGCTTGAGATTTGCGAAGCAAAACCTGATCTGTTGAGCGGGATACGCTTAAGT 531
Db 241 AGTAATATAGATGACATATCAAAATCTTATCAATTTAGTAAGAAAGAAATGAAT 182
OY 532 GATCGAAAAAGAAATTTCTACCAAGAAAGTCAA 566
Db 181 GATGTGAAATTAAGTTGTGATATGAAAAATTTAA 147

RESULT 13
US-09-925-065A-901662/C
; Sequence 901662, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A

```

; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 901662
; LENGTH: 612
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-901662
```

```
Query Match      3.2% Score 36.6; DB 6; Length 612;
Best Local Similarity 52.3%; Pred. No. 7.3;
Matches 81; Conservative 0; Mismatches 74; Indels 0; Gaps 0;
```

```

QY 412 ACGGTTCTCAACAGCCGCAACGCAATGATTACGGCGTTACGGCGATGAAAGAT 471
    |||||
DB 372 AAGATTTCAGAAAATAAATAAATAAAGATTAGATGATATATCAAGAGCTGTTAAT 313
    |||||
QY 472 ACGCTTGAGAAATTTGCCAGCAAACTTGATATCGTTGAGCGGATACAGCTTAAGT 531
    |||||
DB 312 AGTAATATAGATGATGATATCAATATCTTATCAATTTAGTAAAGAAAGAAATGAAT 253
    |||||
QY 532 GATGCGAAAAGAAAATTTGATACGAGAAAGTCAA 566
    |||||
DB 252 GATGCGAAAATTTAGCTGTGATATGAAAAAATTAA 218
    |||||
```

```

RESULT 14
US-09-925-065A-73469/c
; Sequence 73469, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 73469
; LENGTH: 745
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-73469
```

```
Query Match      3.2% Score 36.2; DB 6; Length 745;
Best Local Similarity 56.2%; Pred. No. 10;
Matches 68; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
```

```

QY 250 AGTATTTTCAATATGTTGAACCGATCTTTAGAGCCTCGATTAATGACGAAACTTT 309
    |||||
DB 437 AGTTTTCAGATTTCTTGAATAAATATTTCTTCACTGCTTATGCTGTTAGAGCCTTT 378
    |||||
```

```

QY 310 GAAAACCCGGTAACCTTCAAAAAGATTATPAAACAGACATAAAGATTATCTGCAG 369
    |||||
DB 377 CCAAAACCTGTAATTTTCAAAAATATTTTCACTGCTCATGAGGCGCTGAGATTCAC 318
    |||||
QY 370 A 370
DB 317 A 317
```

```

RESULT 15
US-09-925-065A-73470/c
; Sequence 73470, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 73470
; LENGTH: 745
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-73470
```

```
Query Match      3.2% Score 36.2; DB 6; Length 745;
Best Local Similarity 56.2%; Pred. No. 10;
Matches 68; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
```

```

QY 250 AGTATTTTCAATATGTTGAACCGATCTTTAGAGCCTCGATTAATGACGAAACTTT 309
    |||||
DB 437 AGTTTTCAGATTTCTTGAATAAATATTTCTTCACTGCTTATGCTGTTAGAGCCTTT 378
    |||||
QY 310 GAAAACCCGGTAACCTTCAAAAAGATTATPAAACAGACATAAAGATTATCTGCAG 369
    |||||
DB 377 CCAAAACCTGTAATTTTCAAAAATATTTTCACTGCTCATGAGGCGCTGAGATTCAC 318
    |||||
QY 370 A 370
DB 317 A 317
```

Search completed: February 27, 2006, 11:38:28
Job time : 686.994 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:34:23 ; Search time 268.258 Seconds
(without alignments)
7553.984 Million cell updates/sec

Title: US-10-789-164-3

Perfect score: 1140
Sequence: 1 atgaaaaagaactgagctt.....gtgacaactaaatcctaa 1140

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/1/COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5/COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A/COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B/COMB.seq.*
5: /cgn2_6/ptodata/1/ina/6/COMB.seq.*
6: /cgn2_6/ptodata/1/ina/PCNUS.COMB.seq.*
7: /cgn2_6/ptodata/1/ina/PP.COMB.seq.*
8: /cgn2_6/ptodata/1/ina/RE.COMB.seq.*
9: /cgn2_6/ptodata/1/ina/backfillseq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	55.8	4.9	7218	2	US-08-232-463-14
C 2	54.2	4.8	1101	3	US-09-134-001C-1575
C 3	47.8	4.2	939	3	US-09-902-540-6958
C 4	47.8	4.2	4604	3	US-09-902-540-602
C 5	47.6	4.2	1170	3	US-09-134-001C-1479
C 6	47.6	4.2	3523	3	US-09-710-279-4412
C 7	47.2	4.1	536165	3	US-09-214-808-1
C 8	45	3.9	1116	3	US-09-902-540-3551
C 9	45	3.9	17228	3	US-09-902-540-1170
C 10	42.4	3.7	14554	3	US-09-949-016-15908
C 11	41.8	3.7	3256	3	US-10-104-047-474
C 12	39.4	3.5	1272	3	US-09-902-540-9266
C 13	39.4	3.5	9146	3	US-09-902-540-990
C 14	39	3.4	1785	3	US-09-328-352-1364
C 15	37.4	3.3	378	3	US-09-583-110-789
C 16	37.4	3.3	483	3	US-09-583-110-788
C 17	37.4	3.3	606	3	US-09-107-433-634
C 18	37.2	3.3	1107	3	US-09-902-540-8213
C 19	37.2	3.3	1992	3	US-09-107-532A-3207
C 20	37.2	3.3	12183	3	US-09-902-540-1065
C 21	37.2	3.3	50563	3	US-09-949-016-15821
C 22	36.8	3.2	9048	3	US-08-961-527-159
C 23	36.6	3.2	1141	3	US-09-806-708B-22
C 24	36	3.2	399	3	US-09-621-976-8976

25	35.8	3.1	860	3	US-08-858-207A-172	Sequence 172, App
26	35.6	3.1	1141	3	US-09-806-708B-22	Sequence 22, Appl
27	35.4	3.1	1664976	3	US-08-916-421B-1	Sequence 1, Appl1
28	35.4	3.1	1664976	3	US-09-692-570-1	Sequence 1, Appl1
C 29	35.2	3.1	118382	3	US-09-949-016-15996	Sequence 15996, A
C 30	35.2	3.1	118382	3	US-09-949-016-15997	Sequence 15997, A
C 31	35	3.1	1245	3	US-09-248-796A-1300	Sequence 1300, Ap
C 32	34.6	3.0	4121	3	US-09-638-524A-1	Sequence 1, Appl1
C 33	34.6	3.0	106746	3	US-09-326-402C-1	Sequence 12, Appl1
C 34	34.6	3.0	106746	3	US-09-326-402C-12	Sequence 1, Appl1
C 35	34.4	3.0	4137	3	US-08-832-867-1	Sequence 13300, A
C 36	34.4	3.0	75176	3	US-09-949-016-13300	Sequence 18033, A
C 37	34.2	3.0	474	3	US-09-621-976-18033	Sequence 16661, A
C 38	34.2	3.0	81819	3	US-09-949-016-16661	Sequence 16662, A
C 39	34.2	3.0	81819	3	US-09-949-016-16662	Sequence 2527, Ap
C 40	34	3.0	498	3	US-09-710-279-2527	Sequence 61404, A
C 41	34	3.0	601	3	US-09-949-016-61404	Sequence 89352, A
C 42	34	3.0	601	3	US-09-949-016-89352	Sequence 1348, Ap
C 43	34	3.0	714	3	US-09-134-001C-1348	Sequence 14160, A
C 44	34	3.0	864	3	US-09-270-767-14160	Sequence 3859, Ap
C 45	34	3.0	3504	3	US-09-710-279-3859	

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEITFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSER: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMUN
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-9300
TELEFAX: (703) 683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgpt-F15

US-08-232-463-14

4.9%; Score 55.8; DB 2; Length 7218;

Query Match
Best Local Similarity 3.1%; Pred. No. 4.5e-06;
Matches 12; Conservative 224; Mismatches 151; Indels 0; Gaps 0;

```
QY 197 ATATTATGATGGAGCGCTATGTTGAAAAAGTAACGAGCAAAAGGGGAGACAGTATTT 256
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1446 AGAATTTGTAACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1387
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 257 TTCAATATGTTGAACCGATCTTTAGAGCCTCGGATTAATGACGAGAACTTTGAAAAC 316
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1386 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1327
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 317 CGGTAACTTACAAAGAAATTATAAACAAGCAGATAAAGAGTTATCTGACAGAGATA 376
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1326 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1267
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 377 AGCAATCAGTGAAGTCTTGAAAGATGAAATTTTCAACGCTTCAACAGCCCAACAAC 436
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1266 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1207
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 437 ACGCAATGATTAACGGCTTCAAGGCATGAAAGATACCTTGAGAGATTTGGCAACAA 496
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1206 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1147
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 497 ACCTGATATGCTTGAGCGGATACAGCTTAAGTATGCGCAAAAGAAATTCGTACC 556
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1146 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1087
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 557 AGAAGTCAACGGGGTAAAGATTGCA 583
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1086 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1060
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 2

US-09-134-001C-1575

; Sequence 1575, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NOCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; PRIOR FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674

; SEQ ID NO 1575

; LENGTH: 1101

; TYPE: DNA

; ORGANISM: Staphylococcus epidermidis

US-09-134-001C-1575

Query Match
Best Local Similarity 4.8%; Score 54.2; DB 3; Length 1101;

Matches 229; Conservative 0; Mismatches 248; Indels 6; Gaps 2;

```
QY 465 GAAAGATCGCTTGAGATTGGAGCAAAACCTTGATATCGTTGGAGGGGATACAG 524
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 459 GACACCAATGATGAGAAAATAGATCCAGACGTTATATTTTAAACAGGTAAAG 518
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 525 CTTAAGTATGAGAAAAGAAATTTGTAACGAGAAAGTCAACGGGTAAACGATTGCAAC 584
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 519 TTCAATGCAATTAATAGTAAACCTGTACAACAACATTAAGTAAAGTAAAGTAAAGTAA 578
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 585 GCTTGCTTTTACCAATGCTGTCGGGAAAGTTTCCGGCTTAAAAAGATATGCGCGGCGT 644
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 579 TGTTCATTTACCGATATGATGATTAAGTAACTTAACTTAACTTAACTTAACTTAACTTAA 635
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
QY 645 GCTCCCGCAGATCTGAAATCTTCATCCCTATGATTTAGAAAGCAAAACATGCTGA 704
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 636 AATTAGTTAGATCAGCTATATTTTATCCTTTATATAAAAAATTAAAGGAAATTAATGA 695
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 705 CATTTGTTGTGACATGACACTGGGGCAAGATATGACAAATATCCAAACGCCCA 764
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 696 TTAAGTGTATGATCAATGATGATGAGGAGATACCTAAAGAAAGAAATGATCAACATGA 755
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 765 GCGCAGCTTGACAGACGATGCTGATGCGGGAGCTGACATCATCTGCGGCATCATCC 824
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 756 AAAAGATATGACATGCGTTAGCGAATGCTGTCGAGATGTCATTTATGCTATTAATAC 815
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 825 GCACGCTTGAACCGATTGAAGTATTAACGAAACCGTCAATTTCTACAGCCTCGGCA 884
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 816 --AGTATTAACAAAAGTTGAATTAATTAAGCGAACCGCTATTTTATAGTTAGGTA 872
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 885 CTTTGCTTTGACCAAGCTGACGAGAACAGAACAGACATGCTGTTCACTATCACT 944
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 873 CACAAGCTGATTAATCTTTATCAAAAATCAAGAAAGAAATGATGTACAAACACTG 932
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 945 GAA 947
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 933 GAA 935
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 3

US-09-902-540-6958

; Sequence 6958, Application US/09902540

; Patent No. 6833447

; GENERAL INFORMATION:

; APPLICANT: Goldman, Barry S.

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Wiegand, Roger C.

; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof

; FILE REFERENCE: 38-10(15849)B

; CURRENT APPLICATION NUMBER: US/09/902,540

; PRIOR FILING DATE: 2001-07-10

; PRIOR APPLICATION NUMBER: 60/217,883

; PRIOR FILING DATE: 2000-07-10

; NUMBER OF SEQ ID NOS: 16825

; SEQ ID NO 6958

; LENGTH: 939

; TYPE: DNA

; ORGANISM: Myxococcus xanthus

US-09-902-540-6958

Query Match
Best Local Similarity 4.2%; Score 47.8; DB 3; Length 939;

Matches 109; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

```
QY 697 CATGCTACATTTGTTGTTGTCAGTCACTGGGGCCCAAGATATGACAAATGATCCAAAC 756
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 472 CAGCGGACATGCTCTGCGCCCTACTTCCACTGGGAGCGCAAGCACTTACACCCCGAG 531
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 757 GACCGCAGCGCGAGCTTGCAAGAGCCATGCTGATGCGGAGCTGACATCATGCTGAGC 816
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 532 CCTTACCAAGTCCGCTCGGCCCACTGCGCATGCAACGCGGCGCGCGGTCTGCGC 591
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 817 CATGATCGCAGCTTTAGAAACGATTGAAGTATATAACGAAACCGTCAATTTCTACAGC 876
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 592 GCCACCGCGCAGTCTTCACAGGATGAGCTGTACACAGGCGAAGCGGTGCTTACTCG 651
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 877 CTGCGCACTTGTCTTTGACCAAGGCTGGA 907
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 652 CTGGGAACTTCTGCTTCGGGGGAACTGGA 682
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 4

US-09-902-540-602/c

; Sequence 602, Application US/09902540

; Patent No. 6833447

; GENERAL INFORMATION:

APPLICANT: Goldman, Barry S.
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Wiegand, Roger C.
TITLE OF INVENTION: MYXOCOCCUS xanthus Genome Sequences and Uses Thereof
FILE REFERENCE: 38-10(15849)B
CURRENT APPLICATION NUMBER: US/09/902,540
CURRENT FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: 60/217,883
PRIOR FILING DATE: 2000-07-10
NUMBER OF SEQ ID NOS: 16825
SEQ ID NO 602
LENGTH: 4604
TYPE: DNA
ORGANISM: Myxococcus xanthus
US-09-902-540-602

Query Match 4.2%; Score 47.8; DB 3; Length 4604;
Best Local Similarity 51.7%; Pred. No. 0.001;
Matches 109; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

Qy 697 CATGCTGACATGTTGTTGTGACATGACACAGTGGGCAAGATGACATGATCCAAAC 756
Db 3256 CAGGCGACATGCTCTGCTCCCTACTTCCATGGGAGCGGAGGACCTACACCCCGAG 3197
Qy 757 GACCGCCAGCGCCAGCTTGCAAGACCATGTCTGATGCGGAGCTGACATCATCTCGGC 816
Db 3196 CCTATCCAGAGTCCGCTGCGCCACATGCGCCAGCGGCGCGGCTGTGCTGCGGC 3137
Qy 817 CATCATCCGACGCTTGTAACCGATGTAATATTAAGGAACCGTATTTTCTACAC 876
Db 3136 GCCACCCGACGCTCTCCAGGCGATGAGCTGTACCAAGGAGACCGGTGCTACTCG 3077
Qy 877 CTCGCACTTGTCTTTGACCAAGGCTGA 907
Db 3076 CTGGGAACTTCTCTTCGGGGGAAGTGA 3046

RESULT 5
US-09-134-001C-1479
Sequence 1479; Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 1479
LENGTH: 1170
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1479

Query Match 4.2%; Score 47.6; DB 3; Length 1170;
Best Local Similarity 54.6%; Pred. No. 0.00053;
Matches 95; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

Qy 686 AACGGAAAAACATGCTGACATGTTGTCAGATCACACTGGGGCCAGAGATGACA 745
Db 635 AACGCCAAACAAAGATGATGATCTTTTCAGACATTTGGGTAATGAAGTAAGC 694
Qy 746 ATGATCAAAACGACCGCCAGCGCCAGCTTGCAAGAGCATGTGATGCGGAGCTGACA 805
Db 695 ATCAGCTTAACGACATCACTCAAAAAAATATGCAAAATTTTCGCTGATGACAGGTGTGATG 754
Qy 806 TCATCTGGCCATCATCTGACGACGCTTTGAACCGATTGAAGTATTAACGAA 859

Db 755 TAGTTTAGTACATCCATCATGTGATTCACACAGTAAATGGGTTGACCGTA 808

RESULT 6
US-09-710-279-4412
Sequence 4412; Application US/09710279
Patent No. 6703492
GENERAL INFORMATION:
APPLICANT: KIMBERLY, WILLIAM JOHN
TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
FILE REFERENCE: PU3480US
CURRENT APPLICATION NUMBER: US/09/710,279
CURRENT FILING DATE: 2000-11-09
PRIOR APPLICATION NUMBER: 60/164,258
PRIOR FILING DATE: 1999-11-09
NUMBER OF SEQ ID NOS: 4412
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 4412
LENGTH: 3523
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-09-710-279-4412

Query Match 4.2%; Score 47.6; DB 3; Length 3523;
Best Local Similarity 54.6%; Pred. No. 0.001;
Matches 95; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

Qy 686 AACGGAAAAACATGCTGACATGTTGTTGTGACATGACACTGGGCGCAAGATGACA 745
Db 2635 AACGCAACAAAGATGATGATCTTTTCAGACATTTGGGTAATGAAGTAAGC 2694
Qy 746 ATGATCAAAACGACCGCCAGCGCCAGCTTGCAAGAGCATGTGATGCGGAGCTGACA 805
Db 2695 ATCAGCTTAACGACATCACTCAAAAAAATATGCAAAATTTTCGCTGATGACAGGTGTGATG 2754
Qy 806 TCATCTGGCCATCATCTGACGACGCTTTGAACCGATTGAAGTATTAACGAA 859
Db 2755 TAGTTTAGTACATCCATCATGTGATTCACACAGTAAATGGGTTGACCGTA 2808

RESULT 7
US-09-214-808-1
Sequence 1; Application US/09214808A
Patent No. 6475793
GENERAL INFORMATION:
APPLICANT: Rosenthal, Andre
APPLICANT: Freiberg, Christoph
APPLICANT: Perret, Xavier Philippe
APPLICANT: Broughton, William John
TITLE OF INVENTION: Genomic Sequence of Rhizobium SP. NGR 234 Symbiotic
Patent No. 6475793
FILE REFERENCE: CARP0068
CURRENT APPLICATION NUMBER: US/09/214,808A
CURRENT FILING DATE: 1999-06-22
PRIOR APPLICATION NUMBER: PCT/IB97/00950
PRIOR FILING DATE: 1997-07-10
NUMBER OF SEQ ID NOS: 1
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 536165
TYPE: DNA
ORGANISM: Rhizobium
US-09-214-808-1

Query Match 4.1%; Score 47.2; DB 3; Length 536165;
Best Local Similarity 52.0%; Pred. No. 0.024;
Matches 106; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY	705	CAATGTTGTTGTCACATCACTCGGGGGCCAAAGATATGACATGATCCAAACGACCGCA	764
Db	421592	CAATTTTACAAACACGACACGAGCCGGAAATTGGACCGAAGCCGGCCATTTTGA	421651
QY	765	GCGCCAGCTTTCAGAGCAGCATGTTGATGCGGAGCTGATCATGTCGGCCATCATCC	824
Db	421652	GCAGGCGCTTGTCTCGCAAGCTAATGATGTCGGGAGCGGACGGCTACGTTGGACACGAGCC	421711
QY	825	GCACGCTTAGAAGCCGATGGAATATATACGGAACCGTCATTTTCTACAGCCCTGGCAA	884
Db	421712	GCACCGACCTGCGTGCGATGGAATATACAAACGCCGCCCATCTTCTATTAATGTCGGGAAA	421771
QY	885	CTTTGCTTTGACCAAGGCTGGAC	908
Db	421772	CTTCTTTATGATGACCTCCGAC	421795

RESULT 8
US-09-902-540-3551
; Sequence 3551, Application US/09902540
; Patent No. 6833447

```

: TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
: FILE REFERENCE: 38-10 (15849) B
: CURRENT APPLICATION NUMBER: US/09/902,540
: CURRENT FILING DATE: 2001-07-10
: PRIOR APPLICATION NUMBER: 60/217, 883
: PRIOR FILING DATE: 2000-07-10
: NUMBER OF SEQ ID NOS: 16825
: SEQ ID NO 3551
: LENGTH: 1116
: TYPE: DNA
: ORGANISM: Myxococcus xanthus
: US-09-902-540-3551

```

Query Match	3.9%;	Score 45;	DB 3;	Length 1116
Best Local Similarity	56.4%;	Pred. No. 0.0033;		
Matches	84;	Conservative	0;	Mismatches 65; Indels

Qy	657	CATGCTGACATTTGTTGTCAGATCACTATGGGGCAAGAGATATACATATATCCAAAC	756
Db	457	CAATGTGAGCGCTCATCTGCTGGGGCACTGGGGGACCGAGATCAAGGGGGGAGCCACGG	516
Qy	757	GACCGCCAGCGCCAGCTTGCACAGAGCCATGTCTGATCGCGGAGCTGACATCATCTCTGGC	816
Db	517	CCCGAGAGACCGTAGCTGGGACCGCGCGCTGCTGAGACCGGGGCGCAGAGGGCGCTCATCTGGG	576
Qy	817	CATCATCCGCAGAGCTCTTAGAACCGATTGA	845
Db	577	CACCATCCGACGTGCTCCAGCGCGCTGGA	605

RESULT 9
US-09-902-540-1170/c
; Sequence 1170, Application US/09902540

```

1  APPLICANT: Goldman, Barry S.
2  APPLICANT: Hinkle, Gregory J.
3  APPLICANT: Slater, Steven C.
4  APPLICANT: Wiegand, Roger C.
5  TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
6  FILE REFERENCE: 38-10(15849)B
7  CURRENT APPLICATION NUMBER: US/09/902,540
8  CURRENT FILING DATE: 2001-07-10
9  PRIOR APPLICATION NUMBER: 60/217, 883
10 PRIOR FILING DATE: 2000-07-10
11 NUMBER OF SEQ ID NOS: 16825
12 SEQ ID NO 1170

```

```

;   LENGTH: 17228
;   TYPE: DNA
;   ORGANISM: Myxococcus xanthus
US-09-902-540-1170

```

Query Match 3.9%; Score 45; DB 3; Length 17228;
Best Local Similarity 56.4%; Pred. No. 0.016;
Matches 84; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY	697	CAATGTCACATTTGTTATGTCACAGTACACACTGGGGCCAAAGATATGACACATATATCCAAAC	756
Db	16771	CAATGTGACCGCGTCAATCGTGTGGGCGCACTGGGGGACGAGATACAAAGGGGAGCCACGG	16712
QY	757	GACCGCCACGGCCACGCTTGCAAGACCAATGTCATGATGGGGAGCTGATCATATCGTCGGC	816
Db	16711	CCCGAGGACCGTGAAGCTGGAGCGCGCGCTGTGTGACGCGGGGGCCAAAGGCGGTCAATCGGG	16655

OY 817 CATCATCCGCACGCTTGTAGAACCGATTGA 845
 ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 16651 CACCATCCGCACGCTGTTCAGGCCGCTGGA 16623

RESULT 10
US-09-949-016-15908/c
; Sequence 15908, Application US/09949016

APPLICANT: VENTER, J. Craig et al.

FILE REFERENCE: CL001307

TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

INVENTOR: JAMES W. HARRIS, JR.

ATTORNEY: JAMES W. HARRIS, JR.

;; CURRENT APPLICATION NUMBER: US/09/949,016
;; CURRENT FILING DATE: 2000-04-14
;; PRIOR APPLICATION NUMBER: 60/241,755
;; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237,768
 ;
 ; PRIOR FILING DATE: 2000-10-03
 ;
 ; PRIOR ADDITION NUMBER: 60/231,498

```

; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0

```

```

; SEQ ID NO 15908
; LENGTH: 14554
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15908

```

Query Match	3.7%;	Score 42.4;	DB 3;	Length 14554;
Best Local Similarity	50.5%;	Pred. No. 0.02;		
Matches 103;	Conservative 0;	Mismatches 101;	Indels 0;	Gaps 0;

Oy	AGAGGATTAAGCAATCAGTGAACCTTTGAAGAATTAGAAATTTTCACGGTCTCAACACGC	427
Db	AAAGGAGGAAAAGAAAGAGAAAGAAAAGAAAGAAAGAAAGAAAGAAAGAAAG	8791
Oy	CCAACCAACACGCAATGGATTACCGCGCTCAGGCCATGAAGATAACGCTTGAGAAATTTG	487
Db	AAAGAAAGGAAAGAAAAGAAACCACATAGCCAGACTTGTAAGAAAAGGAAGAAAGAA	8731
Oy	CGAAGCAAACCTTGATATCGTTGGAGCGGATACAGCTTAAGTAGATGCCAAAAAGAAA	547
Db	AAAAGAAAAAAAAGAAAAGAAAAGAACCAAGATAGACCCAAGAGGCGAGTGAATAAGAAAG	8671
Oy	TTTTGTACCAAGAAAGTCAACGGGG	571
Db	CTTCCCTTCAGAAAAGCCAGCCGAG	8647

RESULT 11
US-10-104-047-474
; Sequence 474, Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:

APPLICANT: HELIX RESEARCH INSTITUTE
TITLE OF INVENTION: No. 6943241e1 full length cdna
FILE REFERENCE: H1-A0105
CURRENT APPLICATION NUMBER: US/10/104,047
CURRENT FILING DATE: 2002-03-25
PRIOR APPLICATION NUMBER:
PRIOR FILING DATE:
NUMBER OF SEQ ID NOS: 4096
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 474
LENGTH: 3256
TYPE: DNA
ORGANISM: Homo sapiens
US-10-104-047-474

Query Match 3.7%; Score 41.8; DB 3; Length 3256;
Best Local Similarity 50.2%; Pred. No. 0.059;
Matches 103; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 243 GCGACAGCTATTTTCAATATGTTGACCGATCTTTAGAGCTCGATTATGTACGAG 302
DB 1684 GCGAATATGCTGCATGATGATGAGCAGAGCTGAAAGCTTTATATGAAACGAG 1743
QY 303 AAACCTTGAACCCGCTAACCTATCAAAAGATTATTAACAGAGATAAGATTCGA 362
DB 1744 TCTGTGTGCTTGGGAGAACTCTGAAAGAAAGCTTAACTCTGATGTCAAGGTGGC 1803
QY 363 TCTGACAGCAATTAAGAAATCAAGTGAAGTGAAGATGATGATTTCAAGGTTCGA 422
DB 1804 TGACAGTGTGAAGATTAATCTGTAAGATTTCTTGACAGCAATTAATGATGACAGCCAG 1863
QY 423 CAGCGCCAAACACCAACCAATGAT 447
DB 1864 TGCCACAGAGGAAATAGCAATTGAT 1888

RESULT 12
US-09-902-540-9266
; Sequence 9266, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 9266
; LENGTH: 1272
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-9266

Query Match 3.5%; Score 39.4; DB 3; Length 1272;
Best Local Similarity 59.3%; Pred. No. 0.19; Indels 46; Gaps 0;
Matches 67; Conservative 0; Mismatches 0; Gaps 0;
QY 779 GAGCCATGCTGATGCGGAGCTGACATCATTCGCGCATTCACGACGCTTAGAAC 838
DB 896 GCGCGGTGTGAGCGCGGCGGCGACCTGTCATCGCCATGCGCGCAAGTGTGCGCG 955
QY 839 CGATTGAATATATACGAAACCGTCATTTTCTAGAGCTCGGCACTTTGTC 891
DB 956 GCATGAGTTCTACAAAGGCGGCTCATCGCTACTGCTGGGCAACTTGGCC 1008

RESULT 13
US-09-902-540-990

Sequence 990, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 990
; LENGTH: 9146
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(9146)
; OTHER INFORMATION: unsure at all n locations
US-09-902-540-990

Query Match 3.5%; Score 39.4; DB 3; Length 9146;
Best Local Similarity 59.3%; Pred. No. 0.59;
Matches 67; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 779 GAGCCATGCTGATGCGGAGCTGACATCATGCTGCGCATTCGACGCTTTAGAAC 838
DB 6491 GCGCGGTGTGAGCGCGGCGGCGACCTGTCATGCGCATGCGCGCAAGTGTGCGCG 6550
QY 839 CGATTGAATATATACGAAACCGTCATTTTCTAGAGCTCGGCACTTTGTC 891
DB 6551 GCATGAGTTCTACAAAGGCGGCTCATGCTACTGCTGGGCAACTTGGCC 6603

RESULT 14
US-09-328-352-1364
; Sequence 1364, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; PRIOR FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 1364
; LENGTH: 1785
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-1364

Query Match 3.4%; Score 39; DB 3; Length 1785;
Best Local Similarity 51.4%; Pred. No. 0.31; Indels 85; Gaps 0;
Matches 90; Conservative 0; Mismatches 85; Gaps 0;

QY 375 TAAGATCAGTGAAGTCTTGAAGATGATTTACGGTTCTCAACAGCCCAACAA 434
DB 1005 TAAATAAATCTTGAAGATTTAAAGATATCTTAATATTTGTTTGGCAATTA 1064
QY 435 CCAAGCATGATTAAGCGGCTTCAAGGCAATGAAGATACGCTTGAAGATTGGCAAGCA 494
DB 1065 CCAATTTGAAGATTAAGCGGCTTCAAGGCAATGAAGATACGCTTGAAGATTGGCAAGCA 1124
QY 495 AAACCTGATATGTTGAGCGGATACAGCTTAAGTATGCAAAAAAGAAATT 549
DB 1125 GAGTATTTCTTATATATGCGCTGATGAAACAAAAAGATGTCACAAATTAATT 1179

RESULT 15

```
US-09-583-110-789/c
; Sequence 789, Application US/09583110
; Patent No. 669703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics
; FILE REFERENCE: PATH00-07A
; CURRENT APPLICATION NUMBER: US/09/583,110
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 789
; LENGTH: 378
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-789

Query Match      3.3%; Score 37.4; DB 3; Length 378;
Best Local Similarity 52.2%; Pred. NO. 0.39;
Matches 83; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 688 GCGAAAAACATGCTGACATTGTTGTGTGACATGACACTGGGCCAAGATATGACAAT 747
    |||||
DB 342 GCAGAGAGGAAGCAGATATCCATATATCATGCTCAGATGGGTGTGAGTATCATGATTG 283
    |||||

QY 748 GATCCAAACGACCGCAGCGCCAGCTTGCAAGAGCCATGTCGATCGGAGCTGACATC 807
    |||||
DB 282 GAACCACTGAAGAAACAAAAGCTCTTATATCAAAAGATGATGATTGGGAGCGGATATT 223
    |||||

QY 808 ATCGTGGCCATCATCCGACGCTTTAGAACCGATTGAA 846
    |||||
DB 222 ATCTTGGAGGGCATCTCAGCTTGTGACCATCTGAA 184
    |||||
```

Search completed: February 27, 2006, 06:46:19
Job time : 273.458 secs

QY	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Db	222	GAGCCAAAGTGCAAGTCCGCTCTCAAGAAGCATCTGACCAAGAGCGTCTCGAGAAGCT	1	GATCCAAAGTGCAAGAACTGCTCTCAAGAAGCCGCTCTCAAGAAGCTGCTCAAGAAGCT	58							
RESULT 5	D76107	360 bp mRNA	EST 14-JUN-2005									
LOCUS	D76107	360 bp mRNA	EST 14-JUN-2005									
DEFINITION	CEBK14AD2F Yui Kohara unpublished cDNA:Strain N2 hermaphrodite											
ACCESSION	D76107	embryo Caenorhabditis elegans cDNA clone YK114d2 5', mRNA sequence.										
VERSION	D76107.1	GI:1121891										
KEYWORDS	EST.											
SOURCE	Caenorhabditis elegans											
ORGANISM	Caenorhabditis elegans											
REFERENCE	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;											
AUTHORS	Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.											
TITLE	1 (bases 1 to 360)											
JOURNAL	Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and											
COMMENT	Tabara,H.											
	Toward an expression map of the C.elegans genome											
	Unpublished (1994)											
	Contact: Yui Kohara											
	Genome Biology Lab.											
	National Institute of Genetics											
	Yata 1111, Mishima, Shizuoka 411, Japan											
	Tel: 81-559-81-6854											
	Fax: 81-559-81-6855											
	Email: ykohara@lab.nig.ac.jp.											
FEATURES	Location/Qualifiers											
SOURCE	1..360											
	/organism="Caenorhabditis elegans"											
	/mol_type="mRNA"											
	/strain="N2"											
	/db_xref="taxon:6239"											
	/clone="YK114d2"											
	/sex="hermaphrodite"											
	/dev_stage="embryo"											
	/clone_lib="Yui Kohara unpublished cDNA:Strain N2											
	hermaphrodite embryo"											
ORIGIN												
Query Match	50.5%; Score 30.8; DB 8; Length 360;											
Best Local Similarity	70.7%; Pred. No.45;											
Matches	41; Conservative 0; Mismatches 17; Indels 0; Gaps 0,											
QY	1	GATCCAAAGTGCAAGAACTGCTCTCAAGAAGCCGCTCTCAAGAAGCTGCTCAAGAAGCT	58									
Db	255	GAGCCAAAGTGCAAGTCCGCTCTCAAGAAGCATCTGACCAAGAGCGTCTCGAGAAGCT	312									
RESULT 6	Z14926	395 bp mRNA	EST 19-JUN-1997									
LOCUS	Z14926	395 bp mRNA	EST 19-JUN-1997									
DEFINITION	CELEB2 Chris Martin sorted cDNA library Caenorhabditis elegans cDNA											
ACCESSION	Z14926	clone cm9e2 5', mRNA sequence.										
VERSION	Z14926.1	GI										

(USA) Dept. of Genetics or (UK)
 (USA) Washington Univ. School of Medicine or (UK) MRC Laboratory of
 Molecular Biology
 Box 8233, 4566 Scott Ave., St. Louis, MI 63110, USA, or, Hills
 Road, Cambridge CB2 2QH, UK
 Tel: (USA) (314)3627072 or (UK) (0223)248011
 Fax: (USA) (314)3624137 or (UK) (0223)402008
 Email: twenematecode.wustl.edu or jes@mc-lmba.cambridge.ac.uk
 single read.

FEATURES
 source
 Location/Qualifiers
 1. 395
 /organism="Caenorhabditis elegans"
 /mol_type="mRNA"
 /strain="Bristol N2"
 /db_xref="taxon:6239"
 /clone="cm962"
 /lab_host="MC1061"
 /clone_lib="Chris Martin sorted cDNA library"
 /note="Vector: lambda phage SHX2; Mixed stage
 hemaphrodite cDNA library. Partially normalized by
 successively picking groups of clones that didn't
 hybridize to previously picked clones. Vector: lambda phage
 SHX2 (Lipshitz, D.H. et al., Gene 88:25-36 (1990)) Host:
 MC1061."

ORIGIN
 Query Match 50.5%; Score 30.8; DB 8; Length 395;
 Best local similarity 70.7%; Prid. No. 46;
 Matches 41; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
 QY 1 GATCCAGTCGAGAAACCTGCTCAGAAACCGCTGCTCAGAGAGCTCTCAGAAACT 58
 |||||||
 Db 138 GAGCCAGATCGAATGCTGCTCTCTCAGAGAGCATCTGACCAAGAGACGTGTCAGAGACT 195
 |||||||

RESULT 7
 BU110239
 LOCUS
 DEFINITION
 Caenorhabditis elegans cDNA clone Yk1128f04 5', mRNA, sequence.
 BU110239
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 EST.
 Caenorhabditis elegans
 Caenorhabditis elegans
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 534)
 Kohara, Y., Shin-i, T., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y.
 and Sugano, S.
 A complementary view of the C.elegans genome
 Unpublished (2002)
 Contact: Tadasu Shin-i
 Center For Genetic Resource Information
 National Institute of Genetics
 111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tehin@genes.nig.ac.jp.
 location/Qualifiers
 1. 534
 /organism="Caenorhabditis elegans"
 /mol_type="mRNA"
 /strain="N2"
 /db_xref="taxon:6239"
 /clone="yk1128f04"
 /sex="hermaphrodite"
 /tissue_type="whole animal"
 /dev_stage="L1"
 /clone_lib="unpublished oligo-capped cDNA library, C.
 elegans L1 stage"

FEATURES
 source

ORIGIN

Query Match 50.5%; Score 30.8; DB 3; Length 534;
 Best Local Similarity 70.7%; Pred. No. 48;
 Matches 41; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy 1 GATCCAAGTGAAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAGAACT 58
 |||||
 Db 265 GAGCCAAGTGAAGTGGCTGCTCAAGAAAGCATCTGACCAAGAGAGCTGTCGAGAAGCT 322

RESULT 8
 BJB05202 534 bp mRNA linear EST 26-MAY-2004
 LOCUS BJB05202 unpublished oligo-capped cDNA library, stage I4
 DEFINITION Caenorhabditis elegans cDNA clone yk1547a11 5', mRNA sequence.
 ACCESSION BJB05202
 VERSION BJB05202.1 GI:47708869
 KEYWORDS EST.
 SOURCE Caenorhabditis elegans
 ORGANISM Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
 REFERENCE 1 (bases 1 to 534)
 AUTHORS Kohara,Y., Shin-I,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
 and Sugano,S.
 TITLE A complementary view of the C.elegans genome
 JOURNAL Unpublished (2002)
 COMMENT Contact: Tadasu Shin-i
 Center for Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshini@genes.nig.ac.jp.
 Location/Qualifiers
 1..534
 /organism="Caenorhabditis elegans"
 /mol_type="mRNA"
 /strain="N2"
 /db_xref="taxon:6239"
 /clone="yk1547a11"
 /sex="Hermaphrodite"
 /tissue_type="whole animal"
 /dev_stage="I4"
 /clone_idb="unpublished oligo-capped cDNA library, stage I4"

ORIGIN
 Query Match 50.5%; Score 30.8; DB 3; Length 534;
 Best Local Similarity 70.7%; Pred. No. 48;
 Matches 41; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy 1 GATCCAAGTGAAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAGAACT 58
 |||||
 Db 249 GAGCCAAGTGAAGTGGCTGCTCAAGAAAGCATCTGACCAAGAGAGCTGTCGAGAAGCT 306

RESULT 9
 BJB63981 562 bp mRNA linear EST 25-MAY-2004
 LOCUS BJB63981 unpublished oligo-capped cDNA library Caenorhabditis
 DEFINITION elegans cDNA clone yk1606b07 5', mRNA sequence.
 ACCESSION BJB63981
 VERSION BJB63981.1 GI:47624509
 KEYWORDS EST.
 SOURCE Caenorhabditis elegans
 ORGANISM Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
 REFERENCE 1 (bases 1 to 562)
 AUTHORS Kohara,Y., Shin-I,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
 and Sugano,S.
 TITLE A complementary view of the C.elegans genome
 JOURNAL Unpublished (2002)

COMMENT Contact: Tadasu Shin-i
 Center for Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshini@genes.nig.ac.jp.
 Location/Qualifiers
 1..562
 /organism="Caenorhabditis elegans"
 /mol_type="mRNA"
 /strain="N2"
 /db_xref="taxon:6239"
 /clone="yk1606b07"
 /sex="Hermaphrodite"
 /tissue_type="whole animal"
 /dev_stage="varied"
 /clone_idb="unpublished oligo-capped cDNA library"

ORIGIN
 Query Match 50.5%; Score 30.8; DB 3; Length 562;
 Best Local Similarity 70.7%; Pred. No. 48;
 Matches 41; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy 1 GATCCAAGTGAAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAGAACT 58
 |||||
 Db 246 GAGCCAAGTGAAGTGGCTGCTCAAGAAAGCATCTGACCAAGAGAGCTGTCGAGAAGCT 303

RESULT 10
 BJB68324 581 bp mRNA linear EST 25-MAY-2004
 LOCUS BJB68324 unpublished oligo-capped cDNA library Caenorhabditis
 DEFINITION elegans cDNA clone yk1668h11 5', mRNA sequence.
 ACCESSION BJB68324
 VERSION BJB68324.1 GI:47633175
 KEYWORDS EST.
 SOURCE Caenorhabditis elegans
 ORGANISM Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
 REFERENCE 1 (bases 1 to 581)
 AUTHORS Kohara,Y., Shin-I,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
 and Sugano,S.
 TITLE A complementary view of the C.elegans genome
 JOURNAL Unpublished (2002)
 COMMENT Contact: Tadasu Shin-i
 Center for Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshini@genes.nig.ac.jp.
 Location/Qualifiers
 1..581
 /organism="Caenorhabditis elegans"
 /mol_type="mRNA"
 /strain="N2"
 /db_xref="taxon:6239"
 /clone="yk1668h11"
 /sex="Hermaphrodite"
 /tissue_type="whole animal"
 /dev_stage="varied"
 /clone_idb="unpublished oligo-capped cDNA library"

ORIGIN
 Query Match 50.5%; Score 30.8; DB 3; Length 581;
 Best Local Similarity 70.7%; Pred. No. 49;
 Matches 41; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy 1 GATCCAAGTGAAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAGAACT 58
 |||||
 Db 266 GAGCCAAGTGAAGTGGCTGCTCAAGAAAGCATCTGACCAAGAGAGCTGTCGAGAAGCT 323

[illegible]

	FEATURES	Location/Qualifiers
	source	/l..605 /organism="Caenorhabditis elegans" /mol_type="mRNA" /strain="N2"
		/db_xref="taxon:6239" /clone="YK1213D01" /sex="hermaphrodite" /tissue_type="whole animal" /dev_stage="L1" /clone_lib="unpublished oligo-capped cDNA library, C. elegans L1 stage"
ORIGIN		
Query Match	50.5%; Score 30.8; DB 3; Length 605;	
Best Local Similarity	70.7%; Pred. No. 49;	
Matches	41; Conservative 0; Mismatches 17; Indels 0; Gaps 0;	
Oy	1 GATCCAACTGGAAGAACGTGCTCAAGAAACCGTGCTCACAAGACTGCTCAAGAAGT 58 	
Dn	266 GAGCCAAGTCAGACGTGGCTGCTCAAGAAACATTTGCACCAGACACTGCTCAGAAAGCT 323 	
RESULT 13		
LOCUS	AU200813 609 bp mRNA linear EST 17-JUL-2001	
DEFINITION	AU200813 unpublished oligo-capped cDNA library, stage L1	
ACCESSION	Cenothaditidis elegans cDNA clone YK784c05 5', mRNA sequence.	
VERSION	AU200813	
KEYWORDS	AU200813.1 GI:14829131	
SOURCE	EST.	
ORGANISM	Caenorhabditis elegans Caenorhabditidis elegans Eukaryota; Metazoa; Nematoda; Chromodorea; Rhabditiida; Rhabditoidea; Rhabditiidae; Peloderinae; Caenorhabditis.	
REFERENCE	1 (bases 1 to 609) Kohara, Y., Shin-I, T., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y. and Sugano, S. A complementary view of the C.elegans genome Unpublished (2001)	
JOURNAL COMMENT	Contact: Yuji Kohara Genome Biology Lab. National Institute of Genetics Yata 1111, Mishima, Shizuoka 411, Japan Tel: 81-559-81-6854 Fax: 81-559-81-6855 Email: ykohara@lab.nig.ac.jp.	
FEATURES	location/Qualifiers	
source	1..609 /organism="Caenorhabditis elegans" /mol_type="mRNA" /strain="N2" /db_xref="taxon:6239" /clone="YK784c05" /sex="Hermaphrodite" /tissue_type="Whole animal" /dev_stage="L1" /clone_lib="unpublished oligo-capped cDNA library, stage L1"	
ORIGIN		
Query Match	50.5%; Score 30.8; DB 1; Length 609;	
Best Local Similarity	70.7%; Pred. No. 49;	
Matches	41; Conservative 0; Mismatches 17; Indels 0; Gaps 0;	
Oy	1 GATCCAACTGGAAGAACGTGCTCAAGAAACCGTGCTCACAAGACTGCTCAAGAAGT 58 	
Dn	266 GAGCCAAGTCAGACGTGGCTGCTCAAGAAACATTTGCACCAGACACTGCTCAGAAAGCT 323 	
RESULT 14		
LOCUS	BJ118562 610 bp mRNA linear EST 23-JAN-2002	

DEFINITION	Bu118562 unpublished oligo-capped cDNA library, C. elegans L1 stage
ACCESSION	Bu118562
VERSION	Bu118562.1
KEYWORDS	GI:18278679
SOURCE	EST.
ORGANISM	Caenorhabditis elegans Caenorhabditis elegans Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilita; Rhabdilitoidea; Rhabdilitidae; Peloderinae; Caenorhabditis. 1 (bases 1 to 610)
REFERENCE	Kohars, Y., Shin-1, T., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y. and Sugano, S. A complementary view of the C. elegans genome unpublished (2002)
JOURNAL	Contact: Tadasu Shin-i Center For Genetic Resource Information National Institute of Genetics 111 Yata, Mishima, Shizuoka 411-8540, Japan Tel: 81-559-81-6856 Fax: 81-559-81-6855 Email: tshini@gene.nig.ac.jp.
COMMENT	Location/Qualifiers
FEATURES	1..610
source	/organism="Caenorhabditis elegans" /mol_type="mRNA" /strain="N2" /db_xref="taxon:6239" /clone="yk1231a10" /sex="hermaphrodite" /tissue_type="whole animal" /dev_stage="L1" /clone_lib="unpublished oligo-capped cDNA library, C. elegans L1 stage"
ORIGIN	

RESULT 15	
AU201470	
LOCUS	AU201470 614 bp mRNA linear EST 17-JUL-2001
DEFINITION	AU201470 unpublished oligo-capped cDNA library, stage L1
ACCESSION	Caenorhabditis elegans cDNA clone YK793e08 5', mRNA sequence.
VERSION	AU201470
KEYWORDS	
SOURCE	AU201470.1 GI:14830377
ORGANISM	EST.
	Caenorhabditis elegans
	Caenorhabditis elegans
	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
REFERENCE	Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
AUTHORS	1 (bases 1 to 614)
	Kohara,Y., Shin-I,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
	and Sugano,S.
TITLE	A complementary view of the C.elegans genome
JOURNAL	Unpublished (2001)
COMMENT	Contact: Yuji Kohara
	National Institute of Genetics
	Genome Biology Lab.
	Yata III, Mishima, Shizuoka 411, Japan
	Tel: 81-559-81-6854
	Fax: 81-559-81-6855
	Email: ykohara@lab.nig.ac.jp.
FEATURES	
source	Location/Qualifiers
	1..614
	/organism="Caenorhabditis elegans"
	/mol_type="mRNA"
	/strain="N2"

```

/db_xref="taxon:6239"
/clone="YK793e08"
/sex="Hermaphrodite"
/tissue_type="whole animal"
/dev_stage="L1"
/clone_lib="unpublished oligo-capped cdna library, stage
L1"

ORIGIN

Query Match      50.5%; Score 30.8; DB 1; Length 614;
Best Local Similarity 70.7%; Pred. No. 49;
Matches 41; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Oy      1  GATCCAAGTGAAGAAACTGCTCAAGAACCGCTGCTCAAGAACTGCTCAAGAACT 58
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      264 GAGCCAAGTGAAGTCCGCTGCTCAAGAACCAATTCAGCCAAAGACGTCGTCAGAAAGCT 321

```

```
Oy      1 GATCCAAGTGGAGAALACTGCTCAAGAACCCGTGCTCAAGAAAGCTGCTCAAGAAGT   58
Db      264 GAGCCAAGTGCAGTCGCTGCTCAAGAACATCTGAACCAAGACGTCGTGAGAAAGCT   321
```

```
Search completed: February 27, 2006, 11:09:33
Job time : 348.29 secs
```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:34:16 ; Search time 38.5097 Seconds
(without alignments)
10556.990 Million cell updates/sec

Title: US-10-789-164-4

Perfect score: 61
Sequence: 1 gatccaagtgaagaactg.....aagctgcctcaagaactgta 61

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_21:*

- 1: geneseqn1980s:*
- 2: geneseqn1990s:*
- 3: geneseqn2000s:*
- 4: geneseqn2001as:*
- 5: geneseqn2001bs:*
- 6: geneseqn2002as:*
- 7: geneseqn2002bs:*
- 8: geneseqn2003as:*
- 9: geneseqn2003bs:*
- 10: geneseqn2003cs:*
- 11: geneseqn2003ds:*
- 12: geneseqn2004as:*
- 13: geneseqn2004bs:*
- 14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	61	100.0	61	13	ADV65735	Adv65735 Peptide p
2	46	75.4	62	13	ADV65736	Adv65736 Oligonuc
3	30.4	49.8	882	12	AD130094	Ad130094 M. methyl
4	30.2	49.5	549	8	ADAT0869	Ada70869 Rice gene
5	29.4	48.2	773	3	ACAC3118	Acac3118 Arabidops
6	29.4	48.2	787	6	ABN99086	Abn99086 Arabidops
7	29.4	48.2	827	3	AAC39382	Aac39382 Arabidops
8	29.4	48.2	872	3	AAC51288	Aac51288 Arabidops
9	29.4	48.2	874	3	AAC37192	Aac37192 Arabidops
10	29.4	48.2	1420	3	AAAT78510	Aa78510 Plant SDR
11	29.2	47.9	1372	13	ADSS0247	Adss0247 Bacterial
12	28.6	46.9	296	6	ABL70950	Ab170950 Corn tase
13	28.6	46.9	346	11	ACU33808	Ac133808 Rice abio
14	28.6	46.9	546	12	ADU44427	Adj44427 Plant CDN
15	28.6	46.9	785	11	ACU33278	Ac133278 Rice abio
16	28.6	46.9	785	12	ADU43613	Adj43613 Plant CDN
17	28.6	46.9	866	13	ADK46590	Adk46590 Plant full
18	28.6	46.9	1019	12	ADJ43614	Adj43614 Plant CDN
19	28.2	46.2	1592	2	AAQ14501	Aaq14501 USP45 gen

20	28.2	46.2	1920	2	AAQ13830	Aaq13830 MSP gene
21	28	45.9	916	13	ADS47519	Ads47519 Bacterial
22	27.8	45.6	2103	10	ADC91813	Adc91813 E. faeciu
23	27.8	45.6	160755	4	AAH88704	Aah88704 Human DNA
24	27.6	45.2	3870	4	ABL28047	Ab128047 Drosophi1
25	27.6	45.2	6556	4	ABL28046	Ab128046 Drosophi1
26	27.4	44.9	469	9	ACH14859	Ach14859 Human adu
27	27.4	44.9	1227	4	AA160808	Aa160808 Human pol
28	27.4	44.9	1466	4	AA159022	Aa159022 Human pol
29	27.4	44.9	1466	5	ADO99244	Ado99244 DNA encod
30	27.4	44.9	1466	9	ADB49004	Adb49004 Novel hum
31	27.4	44.9	1473	12	ADP28487	Adp28487 Human sec
32	27.4	44.9	2078	13	ACN43611	Acn43611 Human dia
33	27.4	44.9	2089	13	ACN43610	Acn43610 Human dia
34	27.4	44.9	2093	13	ACN43609	Acn43609 Human dia
35	27.4	44.9	2155	13	ACN43608	Acn43608 Human dia
36	27.4	44.9	2203	13	ACN43607	Acn43607 Human dia
37	27.4	44.9	2207	13	ACN43606	Acn43606 Human dia
38	27.4	44.9	2216	13	ACN43605	Acn43605 Human dia
39	27.4	44.9	2240	13	ACN42417	Acn42417 Human dia
40	27.4	44.9	2241	13	ACN43604	Acn43604 Human dia
41	27.4	44.9	2245	13	ACN43603	Acn43603 Human dia
42	27.4	44.9	2256	12	ADJ27218	Adj27218 Human TRI
43	27.4	44.9	2272	13	ACN43614	Acn43614 Human dia
44	27.4	44.9	2282	13	ACN43613	Acn43613 Human dia
45	27.4	44.9	2300	13	ACN43612	Acn43612 Human dia

ALIGNMENTS

RESULT 1
ADV65735
ID ADV65735 standard; DNA; 61 BP.
XX
AC ADV65735;
XX
DT 10-FEB-2005 (first entry)
XX
DE Peptide p5 encoding DNA SEQ ID NO:4.
XX
KW ds; gene expression; antibiotic; antimicrobial; fungicide; cytostatic;
peptide p5.
XX
OS Synthetic.
XX
PN KR2004034780-A.
XX
PD 29-APR-2004.
XX
XX
XX 17-OCT-2002; 2002KR-00063379.
XX
XX 17-OCT-2002; 2002KR-00063379.
XX
XX 17-OCT-2002; 2002KR-00063379.
XX
PA (BIOL-) BIOLEADERS CORP.
PA (UTCH-) UNIV CHUNSU CO. LTD.
PA (KOR-) KOREA RES INST BIOTECHNOLOGY & BIOTECHNOLOG.
XX
PI Boon HR, Ham GS, Hong SP, Jung CM, Kim CJ, Lee DG, Lee JS;
Park YG, Sung MH;
XX
DR WPI, 2004-577380/56.
XX
XX Method for surface expression of peptides p5 and anal3 using pgs bca
gene.
XX
PS Claim 3; SEQ ID NO 4; 25pp; Korean.
XX
XX The invention relates to a novel method for surface expression of
peptides p5 and Anal3 using a poly-gamma-glutamic acid synthase (pgs)BCA
gene, thereby removing a purification process of peptides p5 and Anal3,
CC and using lactic acid bacteria for the surface expression, so that
CC peptide antibiotics can be cheaply and stably mass-produced. An

CC expression vector pHELIB:pgsa-P5 comprises one or more genes encoding
CC poly-gamma-glutamic acid synthase selected from pgsB, pgsC and pgsA, and
CC a gene for dipolar peptide antibiotics having antimicrobial, antifungal
CC and anticancer activities, wherein the dipolar peptide antibiotic has
CC homology to the peptide p5 encoded by the nucleotide sequence set forth
CC in ADV65735, or to the peptide Anal3 encoded by the nucleotide sequence
CC set forth in ADV65737. The present sequence encodes the peptide p5 used
CC in the invention.

XX Sequence 61 BP; 23 A; 14 C; 14 G; 10 T; 0 U; 0 Other;

SO Query Match 100.0%; Score 61; DB 13; Length 61;
Best Local Similarity 100.0%; Pred. No. 3.2e-10;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATCGAAGTGAAGAACTGCTCAAGAAACCGCTCTCAAGAGCTGCTCAAGAACTGT 60
DB 1 GATCGAAGTGAAGAACTGCTCAAGAAACCGCTCTCAAGAGCTGCTCAAGAACTGT 60
QY 61 A 61
DB 61 A 61

RESULT 2
ADV65736/C
ID ADV65736 standard; DNA; 62 BP.
XX
AC ADV65736;
XX
DT 10-FEB-2005 (first entry)
XX
DE Oligonucleotide of the invention SEQ ID NO:5.
XX
KM ss; gene expression; antibiotic; antimicrobial; fungicide; cytostatic.
XX
OS Synthetic.
XX
PM KR2004034780-A.
XX
PD 29-APR-2004.
XX
PF 17-OCT-2002; 2002KR-00063379.
XX
PR 17-OCT-2002; 2002KR-00063379.
XX
PA (BIOL-) BIOLEADERS CORP.
PA (UYCH-) UNIV CHUNSUN CO LTD.
PA (KORE-) KOREA RES INST BIOSCIENCE & BIOTECHNOLOG.
PI Boon HR, Ham GS, Hong SP, Jung CM, Kim CJ, Lee DG, Lee JS;
PI Park YG, Sung MH;
PI WPI; 2004-577380/56.
XX
PT Method for surface expression of peptides p5 and anal3 using pgs bca
XX gene.
XX
XX Example 1; SEQ ID NO 5; 25pp; Korean.

CC The invention relates to a novel method for surface expression of
CC peptides p5 and Anal3 using a poly-gamma-glutamic acid synthase (pgs)BCA
CC gene, thereby removing a purification process of peptides p5 and Anal3,
CC and using lactic acid bacteria for the surface expression, so that
CC peptide antibiotics can be cheaply and stably mass-produced. An
CC expression vector pHELIB:pgsa-P5 comprises one or more genes encoding
CC poly-gamma-glutamic acid synthase selected from pgsB, pgsC and pgsA, and
CC a gene for dipolar peptide antibiotics having antimicrobial, antifungal
CC and anticancer activities, wherein the dipolar peptide antibiotic has
CC homology to the peptide p5 encoded by the nucleotide sequence set forth
CC in ADV65735, or to the peptide Anal3 encoded by the nucleotide sequence
CC set forth in ADV65737. The present sequence represents an oligonucleotide
CC used in the invention to introduce the peptide p5 encoding DNA into

CC expression vector pHELIB:A-P5.
XX
SO Sequence 62 BP; 11 A; 15 C; 14 G; 22 T; 0 U; 0 Other;

QY Query Match 75.4%; Score 46; DB 13; Length 62;
Best Local Similarity 98.3%; Pred. No. 2.3e-05;
Matches 57; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 5 CAAGTGAAGAACTGCTCAAGAAACCGCTCTCAAGAGCTGCTCAAGAACTGT 61
DB 62 CAAGTGAAGAACTGCTCAAGAAACCGCTCTCAAGAGCTGCTCAAGAACTGT 5

RESULT 3
AD130094
ID AD130094 standard; DNA; 882 BP.
XX
AC AD130094;
XX
DT 22-APR-2004 (first entry)
XX
DE M. methylotrophus D-arabino-3-hexulose 6-phosphate synthase (tps1B) DNA.
XX
KM C1 assimilation; L-amino acid production; fermentation; gene; ds.
XX
OS Methylotrophus methylotrophus.
XX
PM Key Location/Qualifiers
FT CDS 1..882
FT /*tag= a
FT /product= "M. methylotrophus enzyme involved in C1
FT assimilation"
XX
PM US200323238-A1.
XX
PD 18-DEC-2003.
XX
PF 12-JUN-2002; 2002US-00166653.
XX
PR 12-JUN-2002; 2002US-00166653.
XX
PA (USUD/) USUDA Y.
PA (NISH/) NISHIO Y.
PA (YASU/) YASUEDA H.
PA (SUGI/) SUGIMOTO S.
PI Usuda Y, Nishio Y, Yasueda H, Sugimoto S;
PI WPI; 2004-061268/06.
DR P-PSDB; AD130095.
XX
PT New polynucleotides encoding polypeptides involved in one-carbon compound
PT metabolism in Methylotrophus methylotrophus, useful for producing L-amino
XX acids (e.g. threonine or serine) or as probes and/or as primers.
XX
PS Claim 13; SEQ ID NO 11; 53pp; English.

CC The invention relates to enzymes involved in C1 assimilation in
CC Methylotrophus methylotrophus and their corresponding polynucleotides. The
CC invention also relates to methods of producing amino acids in micro-
CC organisms having enhanced or attenuated expression of the polypeptides
CC and polynucleotides involved in C1 assimilation. The invention is useful
CC for producing L-amino acids (e.g. asparagine, threonine or serine) using
CC Methylotrophus bacteria. It is also used in identifying microorganisms
CC that can be used to produce fine chemicals, for e.g. by fermentative
CC processes. The present sequence is M. methylotrophus gene encoding an
XX enzyme involved in C1 assimilation.

SO Sequence 882 BP; 238 A; 204 C; 237 G; 203 T; 0 U; 0 Other;

QY Query Match 49.8%; Score 30.4; DB 12; Length 882;
Best Local Similarity 71.4%; Pred. No. 4.6;
Matches 40; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

OY 6 AAGTGGAGAACTGCTCAAGAAACCGCTGCTCAAGAACTGCTCAAGAACTGTA 61
DB 815 AAGAGGCTGCACTGCTGTAAGCTGCGCTGCTGTAAGAACTGCGCGACGCACTGAA 870

RESULT 4

ADA70869
ID ADA70869 standard; DNA; 549 BP.

ADA70869;

20-NOV-2003 (first entry)

Rice gene, SEQ ID 4192.

Plant; bacterial infection; fungal infection; viral infection; rice;
gene; de.

Oryza sativa.

WO2003000898-A1.

03-JAN-2003.

22-JUN-2001; 2001WO-IB001105.

22-JUN-2001; 2001WO-IB001105.

(SYGN) SYNGENTA PARTICIPATIONS AG.

Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y,
Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;

WPI; 2003-175290/17.

Identifying at least one gene involved in plant resistance or response to
pathogenic infection for conferring resistance or tolerance to a plant to
bacterial, fungal or viral infection by determining or detecting plant
gene expression.

Claim 6; SEQ ID NO 4192; 899bp; English.

The present invention relates to a method (M1) for identifying genes
involved in plant resistance or response to pathogenic infection. M1
comprises identifying a gene whose expression is significantly altered in
the incompatible interaction of plant gene expression relative to
expression of the gene in an uninfected plant, in a mutant plant that
does not express a gene associated with response to pathogenic infection,
or in a corresponding incompatible or compatible interaction. (M1) is
useful for conferring resistance to resistance or tolerance to a plant to
bacterial, fungal or viral infection. The present sequence was used to
illustrate the invention.

Sequence 549 BP; 141 A; 122 C; 171 G; 115 T; 0 U; 0 Other;

Query Match 49.5%; Score 30.2; DB 8; Length 549;

Best Local Similarity 74.5%; Pred. No. 4.8;

Matches 38; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

OY 11 GAAGAACTGCTCAAGAAACCGCTGCTCAAGAACTGCTCAAGAACTGTA 61

DB 205 GATGAACCTGTGAAGAGGAGCTGCTGTAAGCTATTGAAGAACTGGA 255

RESULT 5

AAC35118
ID AAC35118 standard; DNA; 773 BP.

AAC35118;

17-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 9060.

XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
XX promoter; termination sequence; ss.

Arabidopsis thaliana.

EP1033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-00301439.

25-FEB-1999; 99US-0121825P.

05-MAR-1999; 99US-0123180P.

09-MAR-1999; 99US-0123548P.

23-MAR-1999; 99US-0125788P.

25-MAR-1999; 99US-0126264P.

29-MAR-1999; 99US-0126785P.

01-APR-1999; 99US-0127462P.

06-APR-1999; 99US-0128234P.

08-APR-1999; 99US-0128714P.

16-APR-1999; 99US-0129845P.

19-APR-1999; 99US-0130077P.

21-APR-1999; 99US-0130449P.

23-APR-1999; 99US-0130510P.

28-APR-1999; 99US-0130891P.

30-APR-1999; 99US-0131449P.

04-MAY-1999; 99US-0132048P.

05-MAY-1999; 99US-0132407P.

06-MAY-1999; 99US-0132486P.

07-MAY-1999; 99US-0132863P.

11-MAY-1999; 99US-0134256P.

14-MAY-1999; 99US-0134219P.

14-MAY-1999; 99US-0134219P.

14-MAY-1999; 99US-0134219P.

18-MAY-1999; 99US-0134768P.

19-MAY-1999; 99US-0134841P.

20-MAY-1999; 99US-0135124P.

21-MAY-1999; 99US-0135353P.

24-MAY-1999; 99US-0135629P.

25-MAY-1999; 99US-0136021P.

27-MAY-1999; 99US-0136392P.

28-MAY-1999; 99US-0136782P.

01-JUN-1999; 99US-0137222P.

03-JUN-1999; 99US-0137528P.

04-JUN-1999; 99US-0137502P.

07-JUN-1999; 99US-0137724P.

08-JUN-1999; 99US-0138094P.

10-JUN-1999; 99US-0138540P.

10-JUN-1999; 99US-0138847P.

14-JUN-1999; 99US-0139119P.

16-JUN-1999; 99US-0139452P.

16-JUN-1999; 99US-0139453P.

17-JUN-1999; 99US-0139492P.

18-JUN-1999; 99US-0139454P.

18-JUN-1999; 99US-0139455P.

18-JUN-1999; 99US-0139456P.

18-JUN-1999; 99US-0139457P.

18-JUN-1999; 99US-0139458P.

18-JUN-1999; 99US-0139459P.

18-JUN-1999; 99US-0139460P.

18-JUN-1999; 99US-0139461P.

18-JUN-1999; 99US-0139462P.

18-JUN-1999; 99US-0139463P.

18-JUN-1999; 99US-0139750P.

18-JUN-1999; 99US-0139763P.

21-JUN-1999; 99US-0139817P.

PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 03-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 09-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 23-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.

PR 30-AUG-1999; 99US-0151303P.
PR 01-SEP-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152353P.
PR 10-SEP-1999; 99US-0153070P.
PR 15-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157533P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match

48.2%; Score 29.4; DB 3; Length 773;

Best Local Similarity 76.6%; Pred. No. 9.4; Mismatches 11; Indels 0; Gaps 0;

Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 11 GAGAACTGCTCAAGAAACGCTGCTCAAGACCTGCTCAAGAAC 57
|||||
Db 424 GAGAAATATATGACAAACGCTAGTGAAGAACTTGTGAAGAAC 470
|||||

RESULT 6

ABN99086 standard; DNA; 787 BP.

AC ABN99086;

DT 01-AUG-2002 (first entry)

DE Arabidopsis thaliana expressed polynucleotide SEQ ID NO 854.

KW Arabidopsis thaliana; plant; insecticide; fungicide; transgenic; stress;
nutrition; crop; thale cress; tolerance factor; insect; pathogen;

OS Arabidopsis thaliana.

XX US2002023281-A1.
XX
XX 21-FEB-2002.
XX
XX 26-JAN-2001; 2001US-00770445.
XX
XX 27-JAN-2000; 2000US-0178472P.
XX
XX (GORLACH J.
XX (ANYU/ AN Y.
XX (HAMILTON C M.
XX (PRICE J L.
XX (RAINE/ RAINE T M.
XX (YUYU/ YU Y.
XX (RAMEKA J G.
XX (PAGE/ PAGE A.
XX (MATH/ MATH A V.
XX (LEDFORD B L.
XX (WOESSNER J P.
XX (HAAS W D.
XX (GARCIA C A.
XX (KRICKER M.
XX (SLATER T.
XX (DAVIS K R.
XX (ALLEN K.
XX (HOFFMAN N.
XX (HURBAN P.
XX
XX Gorlach J, An Y, Hamilton CM, Price JL, Raines TM, Yu Y;
XX Ramewaka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD;
XX Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N;
XX Hurban P;
XX
XX WPI; 2002-403163/43.
XX
XX New Arabidopsis thaliana nucleic acid for identifying homologous genes,
XX producing compositions that modulate the expression or function of its
XX encoded protein, and mapping functional regions of protein.
XX
XX Claim 1; SEQ ID NO 854; 49pp + Sequence Listing; English.
XX
XX The invention relates to an Arabidopsis thaliana nucleic acid (I)
XX comprising a sequence capable of hybridizing under stringent conditions
XX to a sequence selected from any one of 999 sequences (ABN98233-ABN992331),
XX given in the specification or its fragment. A polypeptide (II) encoded by
XX (I), a transgenic plant (III) comprising an exogenous nucleic acid or a
XX genetically modified cell (IV) comprising an exogenous nucleic acid, is
XX useful for screening a candidate agent for its biological effect. (I) is
XX useful in identifying homologous or related genes, in producing
XX protein, mapping functional regions of the protein and in studying
XX associated physiological pathways. (I) is also useful for the genetic
XX manipulation of cells, particularly plant cells. (I) is also useful in
XX screening assays of various plant strains to determine the strains that
XX are best capable of withstanding a particular disease or environmental
XX stress. (II) and (III) are useful for screening of biologically active
XX agents, e.g. fungicides, insecticides, etc., for elucidating biochemical
XX pathways. The screened agents are useful in improved methods of treating
XX crops to prevent or treat disease. (II) are also useful in screening
XX programs to identify agents that mimic or enhance the action of tolerance
XX factors. Such agents are useful in improved methods of treating crops to
XX enhance their tolerance to environmental stress. (I) is also useful for
XX enhancing or inhibiting production of a biosynthetic product in a plant.
XX (III) is useful for identifying other mediators that may induce
XX expression of proteins of interest, for establishing the extent to which
XX any specific insect and/or pathogen is responsible for damage to a
XX particular plant, for identifying other mediators that enhance or induce
XX tolerance to environmental stress, for identifying factors involved in
XX biosynthetic pathways of nutritional, commercial, or medicinal value and
XX for identifying productions of nutritional, commercial or medicinal
XX value. (IV) is useful in the study of genetic function and regulation,
XX for alteration of the cellular metabolism and for screening compounds

CC that may affect the biological function of the gene or gene products.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?DocID=999909770445
XX
XX SQ Sequence 787 BP; 241 A; 156 C; 173 G; 217 T; 0 U; 0 Other;
XX
XX Query Match 48.2%; Score 29.4; DB 6; Length 787;
XX Best Local Similarity 76.6%; Pred. No. 9.4;
XX Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
XX
XX
XX 11 GAAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAGAAC 57
XX DB 280 GAAGAAATATATGACGAAACCGCTGCTCAAGAAAGTGTCTGAAGAAC 326
XX
XX
XX RESULT 7
XX AAC39382
XX ID AAC39382 standard; DNA; 827 BP.
XX
XX AAC39382;
XX
XX 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 24429.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway; metabolic pathway;
XX promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SBP-2000.
XX
XX
XX 25-FEB-2000; 2000EP-00301439.
XX
XX 25-FEB-1999; 99US-0121825P.
XX 05-MAR-1999; 99US-0123180P.
XX 09-MAR-1999; 99US-0123548P.
XX 23-MAR-1999; 99US-0125788P.
XX 25-MAR-1999; 99US-0126264P.
XX 29-MAR-1999; 99US-0126785P.
XX 01-APR-1999; 99US-0127462P.
XX 06-APR-1999; 99US-0128234P.
XX 08-APR-1999; 99US-0128714P.
XX 16-APR-1999; 99US-0129845P.
XX 19-APR-1999; 99US-0130077P.
XX 21-APR-1999; 99US-0130449P.
XX 23-APR-1999; 99US-0130510P.
XX 28-APR-1999; 99US-0131449P.
XX 30-APR-1999; 99US-0132048P.
XX 30-APR-1999; 99US-0132407P.
XX 04-MAY-1999; 99US-0132484P.
XX 05-MAY-1999; 99US-0132485P.
XX 06-MAY-1999; 99US-0132486P.
XX 06-MAY-1999; 99US-0132487P.
XX 07-MAY-1999; 99US-0132487P.
XX 11-MAY-1999; 99US-0132456P.
XX 14-MAY-1999; 99US-0134218P.
XX 14-MAY-1999; 99US-0134219P.
XX 14-MAY-1999; 99US-0134221P.
XX 14-MAY-1999; 99US-0134370P.
XX 18-MAY-1999; 99US-0134768P.
XX 19-MAY-1999; 99US-0134941P.
XX 20-MAY-1999; 99US-0135124P.
XX 21-MAY-1999; 99US-0135533P.
XX 24-MAY-1999; 99US-0135629P.
XX 25-MAY-1999; 99US-0136021P.
XX 27-MAY-1999; 99US-0136392P.
XX 28-MAY-1999; 99US-0136782P.

PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 28-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.

PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148555P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149729P.
PR 20-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151067P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-015659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 48.2%; Score 29.4; DB 3; Length 827;
Best Local Similarity 76.6%; Pred. No. 9.5;
Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

11 GAAGAACGCTCAAGAACCGCTGCTCAAGAGCTCTCAAGAAC 57
363 GAAGAAATATATGACCAAAACGCTAGTGAAGAGTTGCTGAAGAAC 409

RESULT 8

AAC51288

AAC51288 standard; DNA; 872 BP.

AAC51288;

18-OCT-2000 (first entry)

Arabidopsis thaliana DNA fragment SEQ ID NO: 67963.

Hybridization assay; genetic mapping; gene expression control;
protein identification; signal transduction pathway; metabolic pathway;
promoter; termination sequence; ss.

Arabidopsis thaliana.

EP1033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-00301439.

25-FEB-1999; 99US-0121825P.
05-MAR-1999; 99US-0123180P.
09-MAR-1999; 99US-0123548P.
23-MAR-1999; 99US-0125788P.
25-MAR-1999; 99US-0126264P.
29-MAR-1999; 99US-0126785P.
01-APR-1999; 99US-0128234P.
06-APR-1999; 99US-0128714P.
08-APR-1999; 99US-0129645P.
16-APR-1999; 99US-0130077P.
19-APR-1999; 99US-0130449P.
21-APR-1999; 99US-0130510P.
23-APR-1999; 99US-0130891P.
28-APR-1999; 99US-0131449P.
30-APR-1999; 99US-0132048P.
04-MAY-1999; 99US-0132407P.
04-MAY-1999; 99US-0132484P.
05-MAY-1999; 99US-0132485P.
06-MAY-1999; 99US-0132486P.
07-MAY-1999; 99US-0132487P.
07-MAY-1999; 99US-0132863P.
11-MAY-1999; 99US-0134256P.
14-MAY-1999; 99US-0134218P.
14-MAY-1999; 99US-0134219P.
14-MAY-1999; 99US-0134221P.
14-MAY-1999; 99US-0134370P.
18-MAY-1999; 99US-0134768P.
19-MAY-1999; 99US-0134941P.
20-MAY-1999; 99US-0135124P.
21-MAY-1999; 99US-0135353P.
24-MAY-1999; 99US-0135629P.
25-MAY-1999; 99US-0136021P.
27-MAY-1999; 99US-0136392P.
28-MAY-1999; 99US-0136782P.
01-JUN-1999; 99US-0137222P.
03-JUN-1999; 99US-0137528P.
04-JUN-1999; 99US-0137502P.
07-JUN-1999; 99US-0137724P.
08-JUN-1999; 99US-0138094P.
10-JUN-1999; 99US-0138540P.
10-JUN-1999; 99US-0138847P.

14-JUN-1999; 99US-0139119P.
16-JUN-1999; 99US-0139452P.
16-JUN-1999; 99US-0139453P.
17-JUN-1999; 99US-0139492P.
18-JUN-1999; 99US-0139454P.
18-JUN-1999; 99US-0139455P.
18-JUN-1999; 99US-0139456P.
18-JUN-1999; 99US-0139457P.
18-JUN-1999; 99US-0139458P.
18-JUN-1999; 99US-0139459P.
18-JUN-1999; 99US-0139460P.
18-JUN-1999; 99US-0139461P.
18-JUN-1999; 99US-0139462P.
18-JUN-1999; 99US-0139463P.
18-JUN-1999; 99US-0139750P.
18-JUN-1999; 99US-0139763P.
18-JUN-1999; 99US-0139817P.
21-JUN-1999; 99US-0139899P.
22-JUN-1999; 99US-0140353P.
23-JUN-1999; 99US-0140354P.
24-JUN-1999; 99US-0140695P.
28-JUN-1999; 99US-0140823P.
29-JUN-1999; 99US-0140991P.
30-JUN-1999; 99US-0141287P.
01-JUL-1999; 99US-0141842P.
01-JUL-1999; 99US-0142154P.
02-JUL-1999; 99US-0142055P.
06-JUL-1999; 99US-0142390P.
08-JUL-1999; 99US-0142803P.
09-JUL-1999; 99US-0142920P.
12-JUL-1999; 99US-0142977P.
13-JUL-1999; 99US-0143542P.
14-JUL-1999; 99US-0143624P.
15-JUL-1999; 99US-0144005P.
16-JUL-1999; 99US-0144085P.
16-JUL-1999; 99US-0144086P.
19-JUL-1999; 99US-0144331P.
19-JUL-1999; 99US-0144332P.
19-JUL-1999; 99US-0144333P.
19-JUL-1999; 99US-0144334P.
19-JUL-1999; 99US-0144335P.
19-JUL-1999; 99US-0144352P.
20-JUL-1999; 99US-0144632P.
20-JUL-1999; 99US-0144864P.
21-JUL-1999; 99US-0144814P.
21-JUL-1999; 99US-0145086P.
21-JUL-1999; 99US-0145088P.
22-JUL-1999; 99US-0145087P.
22-JUL-1999; 99US-0145089P.
22-JUL-1999; 99US-0145192P.
22-JUL-1999; 99US-0145145P.
23-JUL-1999; 99US-0145218P.
23-JUL-1999; 99US-0145274P.
26-JUL-1999; 99US-0145276P.
27-JUL-1999; 99US-0145913P.
27-JUL-1999; 99US-0145918P.
27-JUL-1999; 99US-0145919P.
28-JUL-1999; 99US-0146386P.
02-AUG-1999; 99US-0146388P.
02-AUG-1999; 99US-0146389P.
02-AUG-1999; 99US-0146389P.
03-AUG-1999; 99US-0147038P.
04-AUG-1999; 99US-0147204P.
04-AUG-1999; 99US-0147302P.
05-AUG-1999; 99US-0147329P.
05-AUG-1999; 99US-0147260P.
05-AUG-1999; 99US-0147303P.
06-AUG-1999; 99US-0147416P.
09-AUG-1999; 99US-0147493P.
09-AUG-1999; 99US-0147935P.
10-AUG-1999; 99US-0148171P.

```
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159328P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 14-OCT-1999; 99US-0159584P.
PR 18-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.
```

Query Match 48.2%; Score 29.4; DB 3; Length 872;
Best Local Similarity 76.6%; Pred. No. 9.7;
Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

```
OY 11 GAAGAACTGCTCAAGAAACCGCTGCTCAAGAACTGCTCAAGAAAC 57
DB 364 GAAGAAATATGACGAAACCGCTAGTGAAAGATTGCTGAAGAAAC 410
```

```
RESULT 9
AAC37192
ID AAC37192 standard; DNA, 874 BP.
XX
AC AAC37192;
XX
DT 17-OCT-2000 (first entry)
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 16506.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 04-MAY-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134370P.
PR 14-MAY-1999; 99US-0134768P.
PR 18-MAY-1999; 99US-0134941P.
PR 19-MAY-1999; 99US-0135124P.
PR 20-MAY-1999; 99US-0135353P.
PR 21-MAY-1999; 99US-0135629P.
PR 24-MAY-1999; 99US-0136021P.
PR 25-MAY-1999; 99US-0136392P.
PR 27-MAY-1999; 99US-0136782P.
PR 28-MAY-1999; 99US-0137222P.
PR 01-JUN-1999; 99US-0137528P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139411P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
```

PR 18-JUN-1999; 99US-0139457P.
 PR 18-JUN-1999; 99US-0139458P.
 PR 18-JUN-1999; 99US-0139459P.
 PR 18-JUN-1999; 99US-0139460P.
 PR 18-JUN-1999; 99US-0139461P.
 PR 18-JUN-1999; 99US-0139462P.
 PR 18-JUN-1999; 99US-0139463P.
 PR 18-JUN-1999; 99US-0139750P.
 PR 18-JUN-1999; 99US-0139751P.
 PR 21-JUN-1999; 99US-0139817P.
 PR 22-JUN-1999; 99US-0139899P.
 PR 23-JUN-1999; 99US-0140353P.
 PR 24-JUN-1999; 99US-0140354P.
 PR 28-JUN-1999; 99US-0140695P.
 PR 29-JUN-1999; 99US-0140991P.
 PR 30-JUN-1999; 99US-0141287P.
 PR 01-JUL-1999; 99US-0141842P.
 PR 01-JUL-1999; 99US-0141843P.
 PR 02-JUL-1999; 99US-0142055P.
 PR 06-JUL-1999; 99US-0142390P.
 PR 08-JUL-1999; 99US-0142803P.
 PR 09-JUL-1999; 99US-0142920P.
 PR 12-JUL-1999; 99US-0143977P.
 PR 13-JUL-1999; 99US-0143542P.
 PR 14-JUL-1999; 99US-0143624P.
 PR 15-JUL-1999; 99US-0144005P.
 PR 16-JUL-1999; 99US-0144085P.
 PR 16-JUL-1999; 99US-0144086P.
 PR 19-JUL-1999; 99US-0144325P.
 PR 19-JUL-1999; 99US-0144331P.
 PR 19-JUL-1999; 99US-0144332P.
 PR 19-JUL-1999; 99US-0144333P.
 PR 19-JUL-1999; 99US-0144334P.
 PR 19-JUL-1999; 99US-0144335P.
 PR 20-JUL-1999; 99US-0144352P.
 PR 20-JUL-1999; 99US-0144632P.
 PR 20-JUL-1999; 99US-0144684P.
 PR 21-JUL-1999; 99US-0144814P.
 PR 21-JUL-1999; 99US-0145086P.
 PR 21-JUL-1999; 99US-0145088P.
 PR 22-JUL-1999; 99US-0145085P.
 PR 22-JUL-1999; 99US-0145087P.
 PR 22-JUL-1999; 99US-0145089P.
 PR 22-JUL-1999; 99US-0145192P.
 PR 23-JUL-1999; 99US-0145145P.
 PR 23-JUL-1999; 99US-0145218P.
 PR 23-JUL-1999; 99US-0145224P.
 PR 26-JUL-1999; 99US-0145276P.
 PR 27-JUL-1999; 99US-0145913P.
 PR 27-JUL-1999; 99US-0145918P.
 PR 27-JUL-1999; 99US-0145919P.
 PR 28-JUL-1999; 99US-0145951P.
 PR 02-AUG-1999; 99US-0146386P.
 PR 02-AUG-1999; 99US-0146388P.
 PR 03-AUG-1999; 99US-0147038P.
 PR 04-AUG-1999; 99US-0147204P.
 PR 04-AUG-1999; 99US-0147302P.
 PR 05-AUG-1999; 99US-0147192P.
 PR 05-AUG-1999; 99US-0147260P.
 PR 06-AUG-1999; 99US-0147303P.
 PR 06-AUG-1999; 99US-0147416P.
 PR 09-AUG-1999; 99US-0147493P.
 PR 09-AUG-1999; 99US-0147955P.
 PR 10-AUG-1999; 99US-0148171P.
 PR 11-AUG-1999; 99US-0148319P.
 PR 12-AUG-1999; 99US-0148341P.
 PR 13-AUG-1999; 99US-0148565P.
 PR 13-AUG-1999; 99US-0148684P.
 PR 16-AUG-1999; 99US-0149368P.
 PR 17-AUG-1999; 99US-0149175P.
 PR 18-AUG-1999; 99US-0149426P.

PR 20-AUG-1999; 99US-0149722P.
 PR 20-AUG-1999; 99US-0149723P.
 PR 20-AUG-1999; 99US-0149929P.
 PR 23-AUG-1999; 99US-0149902P.
 PR 23-AUG-1999; 99US-0149930P.
 PR 25-AUG-1999; 99US-0150566P.
 PR 26-AUG-1999; 99US-0150884P.
 PR 27-AUG-1999; 99US-0151065P.
 PR 27-AUG-1999; 99US-0151066P.
 PR 27-AUG-1999; 99US-0151080P.
 PR 30-AUG-1999; 99US-0151303P.
 PR 31-AUG-1999; 99US-0151438P.
 PR 01-SEP-1999; 99US-0151930P.
 PR 07-SEP-1999; 99US-0152363P.
 PR 10-SEP-1999; 99US-0153070P.
 PR 13-SEP-1999; 99US-0153758P.
 PR 15-SEP-1999; 99US-0154018P.
 PR 16-SEP-1999; 99US-0154039P.
 PR 20-SEP-1999; 99US-0154779P.
 PR 22-SEP-1999; 99US-0155139P.
 PR 23-SEP-1999; 99US-0155486P.
 PR 24-SEP-1999; 99US-0156559P.
 PR 28-SEP-1999; 99US-0156458P.
 PR 29-SEP-1999; 99US-0156596P.
 PR 04-OCT-1999; 99US-0157117P.
 PR 05-OCT-1999; 99US-0157753P.
 PR 06-OCT-1999; 99US-0157865P.
 PR 07-OCT-1999; 99US-0158029P.
 PR 08-OCT-1999; 99US-0158232P.
 PR 12-OCT-1999; 99US-0158369P.
 PR 13-OCT-1999; 99US-0159293P.
 PR 13-OCT-1999; 99US-0159294P.
 PR 13-OCT-1999; 99US-0159295P.
 PR 14-OCT-1999; 99US-0159329P.
 PR 14-OCT-1999; 99US-0159330P.
 PR 14-OCT-1999; 99US-0159331P.
 PR 14-OCT-1999; 99US-0159637P.
 PR 14-OCT-1999; 99US-0159638P.
 PR 18-OCT-1999; 99US-0159584P.
 PR 21-OCT-1999; 99US-0160741P.
 PR 21-OCT-1999; 99US-0160767P.
 PR 21-OCT-1999; 99US-0160768P.
 PR 21-OCT-1999; 99US-0160770P.
 PR 21-OCT-1999; 99US-0160814P.
 PR 21-OCT-1999; 99US-0160815P.
 PR 22-OCT-1999; 99US-0160980P.
 PR 22-OCT-1999; 99US-0160981P.
 PR 22-OCT-1999; 99US-0160989P.
 PR 25-OCT-1999; 99US-0161404P.
 PR 25-OCT-1999; 99US-0161405P.
 PR 25-OCT-1999; 99US-0161406P.
 PR 26-OCT-1999; 99US-0161359P.
 PR 26-OCT-1999; 99US-0161360P.
 PR 26-OCT-1999; 99US-0161361P.
 PR 28-OCT-1999; 99US-0161920P.
 PR 28-OCT-1999; 99US-0161922P.
 PR 28-OCT-1999; 99US-0161933P.
 PR 29-OCT-1999; 99US-0162142P.

Query Match 48.2%; Score 29.4; DB 3; Length 874;
 Best Local Similarity 76.6%; Pred. No. 9.7;
 Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 11 GAGAACTGCTCAGAAACCGCTGCTCAGAAAGCTGCTCAGAAAC 57
 Db 366 GAGAAATATATGACGAAACGCTAGTGAAGAGTTGCTGAAGAAAC 412

RESULT 10
 ID AAA78510 standard; DNA; 1420 BP.
 XX
 AC AAA78510;

XX 06-AUG-2003 (revised)
DT 27-NOV-2000 (first entry)
XX
DE Plant SDF polynucleotide sequence SEQ List 1 NO:272.
XX
KW Plant; corn; Arabidopsis thaliana; sequence-determined DNA fragment; SDF;
KW genetic mapping; identification; promoter; structural gene; UTR;
KW untranslated region; expression control; ds.
XX
OS Viridiplantae.
XX
PN WO200040695-A2.
XX
PD 13-JUL-2000.
XX
PP 07-JAN-2000; 2000WO-US000466.
XX
PR 08-JAN-1999; 99US-0115293P.
XX
PA (CERE-) CERES INC.
XX
PI Alexandrov N, Brover V, Chen X, Subramanian G, Troukhan ME;
PI Zheng L;
XX
DR WPI; 2000-465970/40.
XX
PT New corn plant and Arabidopsis thaliana sequence-determined DNA
PT fragments, useful for expressing gene products and for controlling
PT expression of a target gene.
XX
PS Claim 1; Page 453-454; 673pp; English.
XX
CC The present invention describes polynucleotides, such as complete cDNA
CC sequences and/or sequences of genomic DNA encompassing complete genes,
CC portions of genes, and/or intergenic regions, collectively referred to as
CC sequence-determined DNA fragments (SDFs), from corn plants and
CC Arabidopsis thaliana. The SDFs are promoters, structural genes,
CC untranslated regions (UTRs), or 3' termination sequences. They can be
CC used for expressing a gene product and controlling expression of a target
CC gene, either as a promoter, a structural gene, an UTR or as a 3'
CC termination sequence. They are also useful as tools for genetic mapping,
CC and identification of a particular individual plant or for clustering a
CC group of plants with a common trait. AAAT8433 to AAAT8630 and AAB24605 to
CC AAAT5099 represent the specifically claimed polynucleotide sequences and
CC polypeptides encoded by them given in the present invention. (Updated on
CC 06-AUG-2003 to correct OS field.)
XX
SQ Sequence 1420 BP; 377 A; 287 C; 348 G; 408 T; 0 U; 0 Other;
XX
Query Match 48.2%; Score 29.4; DB 3; Length 1420;
Best Local Similarity 76.6%; Pred. No. 11;
Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
XX
QY 11 GAGAAACTGCTCAAGAAACCGCTGCTCAGAGAGCTGCTCAAGAAAC 57
DB 359 GAAAGAAATTAATGACGAAACCGCTAGTGAAGAGTTGCTGAAGAAAC 405
XX
RESULT 11
ID ADS50247 standard; cDNA; 1372 BP.
XX
AC ADS50247;
XX
DT 02-DEC-2004 (first entry)
XX
DE Bacterial polynucleotide #4990.
XX
KW Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;

KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW bacterial polynucleotide; gene; ss.
XX
OS Bacteria.
XX
PN US2003233675-A1.
XX
PD 18-DEC-2003.
XX
PP 20-FEB-2003; 2003US-00369493.
XX
PR 21-FEB-2002; 2002US-0360039P.
XX
PA (CAOY/) CAO Y.
PA (HINKLE/) HINKLE G J.
PA (SLATER/) SLATER S C.
PA (CHEN/) CHEN X.
PA (GOLD/) GOLDMAN B S.
XX
PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX
DR WPI; 2004-061375/06.
XX
PT New recombinant DNA construct comprising a promoter positioned to provide
PT for expression of a polynucleotide encoding a polypeptide from a
PT microbial source, useful for producing plants with improved properties.
XX
PS Claim 1; SEQ ID NO 28677; 122pp; English.
XX
CC The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polynucleotide used in
CC the scope of the invention. Note: The sequence data for this patent did
CC not form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 1372 BP; 393 A; 312 C; 322 G; 345 T; 0 U; 0 Other;
XX
Query Match 47.9%; Score 29.2; DB 13; Length 1372;
Best Local Similarity 69.0%; Pred. No. 12;
Matches 40; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
XX
QY 1 GATCCAAGTGAAGAAACTGCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAGAACT 58
DB 547 GAGCCAAGTGAAGACTCGCTGCTCAAGAAAGCATTTGACCAAGAGAGTGTGAGAGCT 604
XX
RESULT 12
ID ABL70950 standard; cDNA; 296 BP.
XX
AC ABL70950;
XX
DT 14-MAY-2002 (first entry)

[illegible]

```

XX OS Eukaryota.
XX PN US2004016025-A1.
XX PD 22-JAN-2004.
XX PF 26-SEP-2002; 2002US-00260238.
XX PR 26-SEP-2001; 2001US-0325277P.
XX PR 26-SEP-2001; 2001US-0325448P.
XX PR 04-APR-2002; 2002US-0370620P.
XX PA (BUDW/) BUDWORTH P.
XX PA (MOUG/) MOUGHAMER T.
XX PA (BRIG/) BRIGGS S P.
XX PA (COOP/) COOPER B.
XX PA (GLAZ/) GLAZEBROOK J.
XX PA (GOFF/) GOFF S A.
XX PA (KATA/) KATAGIRI F.
XX PA (KREP/) KREPS J.
XX PA (PROV/) PROVART N.
XX PA (RICK/) RICE D.
XX PA (ZHUT/) ZHU T.
XX PI Budworth P, Moughamer T, Briggs SP, Cooper B, Glazebrook J;
XX PI Goff SA, Katagiri F, Kreps J, Provart N, Ricke D, Zhu T;
XX DR WPI; 2004-190374/18.
XX PT New rice promoter, useful for manipulating crop plants to alter or
XX PT improve phenotypic characteristics, e.g. produce large quantities of oil
XX PT or proteins, resistance to insecticides, virus or fungi, stress tolerance
XX PT or high nutritional value.
XX PS Example 13; SEQ ID NO 5427; 230pp; English.
XX SC The invention relates to plant nucleotide sequences that direct seed-,
XX SC leaf- and/or stem-, panicle-, root- or pollen-specific or -preferential
XX SC or constitutive transcription of an operatively linked nucleic acid
XX SC segment. The invention also relates to a method for augmenting a plant
XX SC genome and a method of identifying a gene, where its expression is
XX SC altered in the seed, leaf, stem, panicle, pollen, root or is constitutive
XX SC in a plant cell. The plant is a cereal, e.g. soybean, alfalfa, sunflower,
XX SC canola, cotton, peanut, tobacco or sugar beet, preferably maize, barley,
XX SC sorghum, rice or wheat. The polynucleotides and the polypeptides they
XX SC encode are useful for manipulating crop plants to alter or improve
XX SC phenotypic characteristics, to produce large quantities of oil or
XX SC proteins, to incur resistance to insecticides, viruses or fungi, and to
XX SC incur stress tolerance (e.g. salt, cold or drought) to ensure the plants
XX SC have a high nutritional value with reduced apical dominance or dwarfism,
XX SC early flowering or altered metabolic pathways. This sequence represents a
XX SC plant nucleic acid of the invention. Note: The sequence data for this
XX SC patent did not form part of the printed specification but was obtained in
XX SC electronic format directly from USPTO at seqdata.uspto.gov/sequence.html.
XX SQ Sequence 546 BP; 143 A; 123 C; 169 G; 111 T; 0 U; 0 Other.
XX
XX Query Match 46.9%; Score 28.6; DB 12; Length 546;
XX Best Local Similarity 72.5%; Pred. No. 16;
XX Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
XX
XX 11 GAAGAAAGTCTCAAGAAACCGCTGCTCAAGAAAGTCTCAAGAAAGTGA 61
XX 205 GATGAAAGCTTTGAAGAGCAGCAGCTCGAAGAACTATGAAGAAGCTGGA 255
XX
XX RESULT 15
XX ACL33278
XX ID ACJ33278 standard; cDNA; 785 BP.
XX ACJ33278;
XX

```

```

DT 02-JUN-2005 (first entry)
XX DE Rice abiotic stress response related polynucleotide SEQ ID NO:11841.
XX KM ss; abiotic stress tolerance; transgenic plant; plant; cereal;
XX KM agriculture.
XX OS Oryza sativa.
XX PN WO2003008540-A2.
XX PD 30-JAN-2003.
XX PF 21-JUN-2002; 2002WO-US019668.
XX PR 22-JUN-2001; 2001US-0300112P.
XX PR 24-AUG-2001; 2001US-0314662P.
XX PR 26-SEP-2001; 2001US-0325277P.
XX PR 21-NOV-2001; 2001US-0332132P.
XX PA (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX PI Kreps J, Briggs SP, Cooper B, Glazebrook J, Goff SA, Katagiri F;
XX PI Moughamer T, Provart N, Ricke D, Zhu T;
XX DR WPI; 2003-248011/24.
XX PT New stress-responsive nucleic acid, useful for altering the
XX PT responsiveness of a plant, e.g. cereal, to an abiotic stress such as cold
XX PT stress, salt stress or osmotic stress.
XX PS Claim 48; SEQ ID NO 11841; 89pp; English.
XX SC The invention relates to novel abiotic stress responsive polynucleotides
XX SC and polypeptides. Also disclosed are vectors, expression cassettes, host
XX SC cells, and plants containing such polynucleotides. Also disclosed are
XX SC methods for using the polynucleotides and polypeptides to alter the
XX SC responsiveness of a plant to abiotic stress. The invention is useful in
XX SC agriculture. The nucleic acid is useful for determining whether a test
XX SC plant has been exposed to an abiotic stress condition. It is also useful
XX SC for selecting an agent that alters abiotic stress regulated
XX SC polynucleotide expression in a plant cell, and to identify a homolog or
XX SC ortholog to an abiotic stress responsive polynucleotide. The nucleic acid
XX SC molecule and the polypeptide encoded by it are useful in altering the
XX SC responsiveness of a plant to an abiotic stress, such as cold stress, salt
XX SC stress, osmotic stress or any of their combinations. The present sequence
XX SC is used in the exemplification of the invention
XX SQ Sequence 785 BP; 183 A; 198 C; 241 G; 163 T; 0 U; 0 Other;
XX
XX Query Match 46.9%; Score 28.6; DB 11; Length 785;
XX Best Local Similarity 72.5%; Pred. No. 17;
XX Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
XX
XX 11 GAAGAAAGTCTCAAGAAACCGCTGCTCAAGAAAGTCTCAAGAAAGTGA 61
XX 246 GATGAAAGCTTTGAAGAGCAGCAGCTCGAAGAACTATGAAGAAGCTGGA 296
XX

```

Search completed: February 27, 2006, 07:05:24
Job time : 40.7097 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:34:28 ; Search time 66.7965 Seconds
(without alignments)
7551.777 Million cell updates/sec

Title: US-10-789-164-4

Perfect score: 61 gatccaagtgcgaagaactg.....aagctgcctcaagaactgta 61

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Published Applications NA Main:

1: /cgn2_6/pcodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/pcodata/1/pubpna/US08_PUBCOMB.seq.*
3: /cgn2_6/pcodata/1/pubpna/US09_PUBCOMB.seq.*
4: /cgn2_6/pcodata/1/pubpna/US09B_PUBCOMB.seq.*
5: /cgn2_6/pcodata/1/pubpna/US10A_PUBCOMB.seq.*
6: /cgn2_6/pcodata/1/pubpna/US10B_PUBCOMB.seq.*
7: /cgn2_6/pcodata/1/pubpna/US10C_PUBCOMB.seq.*
8: /cgn2_6/pcodata/1/pubpna/US10D_PUBCOMB.seq.*
9: /cgn2_6/pcodata/1/pubpna/US10E_PUBCOMB.seq.*
10: /cgn2_6/pcodata/1/pubpna/US11_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	61	100.0	61	9 US-10-789-164-4	Sequence 4, Appl
2	46	75.4	62	9 US-10-789-164-5	Sequence 5, Appl
3	30.4	49.8	882	6 US-10-166-653-11	Sequence 11, Appl
4	30.4	49.8	882	10 US-11-116-192-11	Sequence 11, Appl
5	30.2	49.5	1254	7 US-10-437-963-46191	Sequence 46191, A
6	29.6	48.5	1717	7 US-10-437-963-22178	Sequence 22178, A
7	29.4	48.2	787	3 US-09-770-445-854	Sequence 854, App
8	29.2	47.9	1372	6 US-10-369-493-28677	Sequence 28677, A
9	28.6	46.9	296	3 US-09-294-0938-324	Sequence 324, App
10	28.6	46.9	546	7 US-10-260-238-5427	Sequence 5427, App
11	28.6	46.9	785	7 US-10-260-238-4613	Sequence 4613, App
12	28.6	46.9	866	7 US-10-425-114-21310	Sequence 21310, A
13	28.6	46.9	1019	7 US-10-260-238-4614	Sequence 4614, App
14	28.6	46.9	1166	8 US-10-425-115-59277	Sequence 59277, A
15	28.6	46.9	1197	8 US-10-425-115-59276	Sequence 59276, A
16	28.6	46.9	916	6 US-10-369-493-25949	Sequence 25949, A
17	27.8	45.6	588	6 US-09-925-065A-246263	Sequence 246263, A
18	27.6	45.2	3870	10 US-11-097-143-39311	Sequence 39311, A
19	27.6	45.2	6556	10 US-11-097-143-39311	Sequence 39311, A
20	27.4	44.9	469	3 US-09-918-995-2071	Sequence 2071, App
21	27.4	44.9	1466	5 US-10-037-370-914	Sequence 914, App
22	27.4	44.9	1466	6 US-10-117-722-914	Sequence 914, App
23	27.4	44.9	1466	9 US-10-122-851-914	Sequence 914, App

24	27.4	44.9	2253	3 US-09-795-693-15	Sequence 15, Appl
25	27.4	44.9	2253	5 US-10-156-239-15	Sequence 15, Appl
26	27.4	44.9	2253	5 US-10-199-485-15	Sequence 15, Appl
27	27.4	44.9	2317	6 US-10-094-749-801	Sequence 801, App
28	27.4	44.9	2593	3 US-09-795-693-13	Sequence 13, Appl
29	27.4	44.9	2593	3 US-10-156-239-13	Sequence 13, Appl
30	27.4	44.9	2593	5 US-10-199-485-13	Sequence 13, Appl
31	27.4	44.9	9733	5 US-09-764-891-6690	Sequence 6890, App
32	27.4	44.9	9738	3 US-09-764-891-6691	Sequence 6891, App
33	27.4	44.9	12348	7 US-10-038-854-3	Sequence 3, Appl
34	27.4	44.9	13935	9 US-10-450-763-10387	Sequence 10387, A
35	27.2	44.6	1057	7 US-10-425-114-24607	Sequence 24607, A
36	27.2	44.6	1146	8 US-10-425-115-59279	Sequence 59279, A
37	27	44.3	369	7 US-10-260-238-1163	Sequence 1163, App
38	27	44.3	423	7 US-10-437-963-13595	Sequence 13595, A
39	27	44.3	499	7 US-10-437-963-35811	Sequence 35811, A
40	27	44.3	588	8 US-10-425-115-4089	Sequence 4089, App
41	27	44.3	1122	7 US-10-437-963-13594	Sequence 13594, A
42	27	44.3	1518	7 US-10-437-963-35812	Sequence 35812, A
43	27	44.3	1620	6 US-10-369-493-33104	Sequence 33104, A
44	27	44.3	1707	6 US-10-369-493-32847	Sequence 32847, A
45	27	44.3	2016	6 US-10-369-493-41284	Sequence 41284, A

ALIGNMENTS

```

RESULT 1
US-10-789-164-4
; Sequence 4, Application US/10789164
; Publication No. US20050191720A1
; GENERAL INFORMATION:
; APPLICANT: Sung, M.H. et al.
; TITLE OF INVENTION: SURFACE EXPRESSION METHOD OF PEPTIDES P5 AND ANAL3 USING THE GENE
; FILE REFERENCE: P1574
; CURRENT APPLICATION NUMBER: US/10/789,164
; CURRENT FILING DATE: 2004-02-27
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Kopatentim 1.71
; LENGTH: 61
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: primer
US-10-789-164-4

Query Match      100.0%; Score 61; DB 9; Length 61;
Best Local Similarity 100.0%; Pred No. 5e-10; 0; Indels 0; Gaps 0;
Matches 61; Conservative 0; Mismatches 0;

QY      1 GATCCAAGTGAAGAACTGCTCAAGAAACCGTGTCTCAAGAACTGCTCAAGAACTGT 60
DB      1 GATCCAAGTGAAGAACTGCTCAAGAAACCGTGTCTCAAGAACTGCTCAAGAACTGT 60

QY      61 A 61
DB      61 A 61

RESULT 2
US-10-789-164-5/c
; Sequence 5, Application US/10789164
; Publication No. US20050191720A1
; GENERAL INFORMATION:
; APPLICANT: Sung, M.H. et al.
; TITLE OF INVENTION: SURFACE EXPRESSION METHOD OF PEPTIDES P5 AND ANAL3 USING THE GENE
; FILE REFERENCE: P1574
; CURRENT APPLICATION NUMBER: US/10/789,164
; CURRENT FILING DATE: 2004-02-27
; NUMBER OF SEQ ID NOS: 9

```

```
SOFTWARE: Koparentin 1.71
; SEQ ID NO 5
; LENGTH: 62
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: primer
US-10-789-164-5
```

```
Query Match          75.4%; Score 46; DB 9; Length 62;
Best Local Similarity 98.3%; Pred. No. 3.3e-05;
Matches 57; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
```

```
OY 5 CAAGTGAAGAACTGCTCAAGAAACC-GCTGCTCAAGAGCTGCTCAAGAACTGTA 61
DB 62 CAAGTGAAGAACTGCTCAAGAAACC-GCTGCTCAAGAGCTGCTCAAGAACTGTA 5
```

RESULT 3

```
US-10-166-653-11
; Sequence 11, Application US/10166653
; Publication No. US20030232338A1
; GENERAL INFORMATION:
```

```
APPLICANT: USUDA, Yoshihiro
APPLICANT: NISHIO, Yosuke
APPLICANT: YASUEDA, Hisaaki
APPLICANT: SUGIMOTO, Shintichi
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING POLYPEPTIDES INVOLVED IN ONE-CARBON COME
```

```
FILE REFERENCE: 211826USO
CURRENT APPLICATION NUMBER: US/10/166,653
```

```
CURRENT FILING DATE: 2002-06-12
```

```
NUMBER OF SEQ ID NOS: 40
```

```
SOFTWARE: PatentIn version 3.1
```

```
SEQ ID NO 11
```

```
LENGTH: 882
```

```
TYPE: DNA
```

```
ORGANISM: Methylophilus methylotrophus
```

```
FEATURE:
```

```
NAME/KEY: CDS
```

```
LOCATION: (1)..(882)
```

```
OTHER INFORMATION:
```

```
US-10-166-653-11
```

```
Query Match          49.8%; Score 30.4; DB 6; Length 882;
Best Local Similarity 71.4%; Pred. No. 6.5;
```

```
Matches 40; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
```

```
OY 6 AAGTGAAGAACTGCTCAAGAAACC-GCTGCTCAAGAGCTGCTCAAGAACTGTA 61
DB 815 AAGAGGCTGCACCTGCTCAAGAGCTGCGCTGCTGAAGAGCTGCGCCAGCACTGAA 870
```

RESULT 4

```
US-11-116-192-11
```

```
; Sequence 11, Application US/11116192
```

```
; Publication No. US20050208634A1
```

```
; GENERAL INFORMATION:
```

```
APPLICANT: USUDA, Yoshihiro
```

```
APPLICANT: NISHIO, Yosuke
```

```
APPLICANT: YASUEDA, Hisaaki
```

```
APPLICANT: SUGIMOTO, Shintichi
```

```
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING POLYPEPTIDES INVOLVED IN ONE-CARBON COME
```

```
FILE REFERENCE: 211826USO
```

```
CURRENT APPLICATION NUMBER: US/11/116,192
```

```
CURRENT FILING DATE: 2005-04-28
```

```
PRIOR APPLICATION NUMBER: US/10/166,653
```

```
PRIOR FILING DATE: 2002-06-12
```

```
NUMBER OF SEQ ID NOS: 40
```

```
SOFTWARE: PatentIn version 3.1
```

```
SEQ ID NO 11
```

```
LENGTH: 882
```

```
TYPE: DNA
; ORGANISM: Methylophilus methylotrophus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(882)
; OTHER INFORMATION:
```

```
US-11-116-192-11
```

```
Query Match          49.8%; Score 30.4; DB 10; Length 882;
Best Local Similarity 71.4%; Pred. No. 6.5;
```

```
Matches 40; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
```

```
OY 6 AAGTGAAGAACTGCTCAAGAAACC-GCTGCTCAAGAGCTGCTCAAGAACTGTA 61
DB 815 AAGAGGCTGCACCTGCTCAAGAGCTGCGCTGCTGAAGAGCTGCGCCAGCACTGAA 870
```

RESULT 5

```
US-10-437-963-46191
```

```
; Sequence 46191, Application US/10437963
```

```
; Publication No. US20040123343A1
```

```
; GENERAL INFORMATION:
```

```
APPLICANT: La Rosa, Thomas J.
```

```
APPLICANT: Kovalic, David K.
```

```
APPLICANT: Zhou, Yihua
```

```
APPLICANT: Cao, Yongwei
```

```
APPLICANT: Wu, Wei
```

```
APPLICANT: Boukharov, Andrey A.
```

```
APPLICANT: Barbazuk, Brad
```

```
APPLICANT: Li, Ping
```

```
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
```

```
FILE REFERENCE: 38-21(53221)B
```

```
CURRENT APPLICATION NUMBER: US/10/437,963
```

```
CURRENT FILING DATE: 2003-05-14
```

```
NUMBER OF SEQ ID NOS: 204966
```

```
SEQ ID NO 46191
```

```
LENGTH: 1254
```

```
TYPE: DNA
```

```
ORGANISM: Oryza sativa
```

```
FEATURE:
```

```
OTHER INFORMATION: Clone ID: PAT_MRT4530_49081C.1
```

```
US-10-437-963-46191
```

```
Query Match          49.5%; Score 30.2; DB 7; Length 1254;
Best Local Similarity 74.5%; Pred. No. 8.2;
```

```
Matches 38; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
```

```
OY 11 GAAGAACTGCTCAAGAAACC-GCTGCTCAAGAGCTGCTCAAGAACTGTA 61
DB 443 GATGAACCTGTTGAAGAGCGACGCTGCTGTGAAGCTATTGAAGAGCTGGA 493
```

RESULT 6

```
US-10-437-963-22178/c
```

```
; Sequence 22178, Application US/10437963
```

```
; Publication No. US20040123343A1
```

```
; GENERAL INFORMATION:
```

```
APPLICANT: La Rosa, Thomas J.
```

```
APPLICANT: Kovalic, David K.
```

```
APPLICANT: Zhou, Yihua
```

```
APPLICANT: Cao, Yongwei
```

```
APPLICANT: Wu, Wei
```

```
APPLICANT: Boukharov, Andrey A.
```

```
APPLICANT: Barbazuk, Brad
```

```
APPLICANT: Li, Ping
```

```
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
```

```
FILE REFERENCE: 38-21(53221)B
```

```
CURRENT APPLICATION NUMBER: US/10/437,963
```

```
CURRENT FILING DATE: 2003-05-14
```

```
NUMBER OF SEQ ID NOS: 204966
```

```
SEQ ID NO 22178
```


LENGTH: 1717
TYPE: DNA
ORGANISM: Oryza sativa
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT4530_27379C.1
US-10-437-963-22178

Query Match 48.5%; Score 29.6; DB 7; Length 1717;
Best Local Similarity 73.1%; Pred. No. 14;
Matches 38; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 4 CCAGTGAAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAGAA 55
DB 502 CAAGAAAGAGAGAGAGCTCAAGAAAGCGCTGCCCAAGAGCAGCTGAGAAA 451

RESULT 7
US-09-770-445-854

Sequence 854, Application US/09770445
Patent No. US20020023281A1
GENERAL INFORMATION:
APPLICANT: Goriach, Jörn
APPLICANT: An, Yong-Oiang
APPLICANT: Hamilton, Carol M.
APPLICANT: Price, Jennifer L.
APPLICANT: Raines, Tracy M.
APPLICANT: Yu, Yang
APPLICANT: Rameaka, Joshua G.
APPLICANT: Page, Amy
APPLICANT: Mathew, Abraham V.
APPLICANT: Ledford, Brooke L.
APPLICANT: Woessner, Jeffrey P.
APPLICANT: Haas, William David
APPLICANT: Garcia, Carlos A.
APPLICANT: Krieker, Maja
APPLICANT: Slader, Ted
APPLICANT: Davis, Keith R.
APPLICANT: Allen, Keith
APPLICANT: Hoffman, Neil
APPLICANT: Hurban, Patrick
TITLE OF INVENTION: Expressed Sequences of Arabidopsis
FILE REFERENCE: 2023US (PAPA-012PRV)
CURRENT APPLICATION NUMBER: US/09/770,445
CURRENT FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: US 60/178,472
PRIOR FILING DATE: 2000-01-27
NUMBER OF SEQ ID NOS: 999
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 854
LENGTH: 787
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-770-445-854

Query Match 48.2%; Score 29.4; DB 3; Length 787;
Best Local Similarity 76.6%; Pred. No. 13;
Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 11 GAAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAGAAAC 57
DB 280 GAAGAAATATATGACGAAACGCTAGTGAAGAGTGTGAGAAAC 326

RESULT 8
US-10-369-493-28677
Sequence 28677, Application US/10369493
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 28677
LENGTH: 1372
TYPE: DNA
ORGANISM: Caenorhabditis elegans
US-10-369-493-28677

Query Match 47.9%; Score 29.2; DB 6; Length 1372;
Best Local Similarity 69.0%; Pred. No. 18;
Matches 40; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 1 GATCCAGTGAAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAGAACT 58
DB 547 GAGCCAGTGAAGAGTGGCTGCTCAAGAAAGCATTTGACCAAGAGTGTGCGAAGACT 604

RESULT 9
US-09-294-093B-324

Sequence 324, Application US/09294093B
Patent No. US20010051335A1
GENERAL INFORMATION:
APPLICANT: Laljudi, Raghunath, V.
APPLICANT: Ito, Laura, Y.
APPLICANT: Sherman, Bradley, K.
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL
FILE REFERENCE: PL-0009 US
CURRENT APPLICATION NUMBER: US/09/294,093B
CURRENT FILING DATE: 1999-04-16
PRIOR APPLICATION NUMBER: 60/082,567
PRIOR FILING DATE: April 21, 1998
NUMBER OF SEQ ID NOS: 6207
SOFTWARE: PERL Program
SEQ ID NO 324
LENGTH: 296
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
NAME/KEY: misc. feature
OTHER INFORMATION: Incyte ID No. US20010051335A1 700342556H1
US-09-294-093B-324

Query Match 46.9%; Score 28.6; DB 3; Length 296;
Best Local Similarity 72.5%; Pred. No. 19;
Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 11 GAAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAGAACTGTA 61
DB 16 GATGAAGCTTTGAAGAGGAGGAGCAGCAGCTCGAAGAGCTATATGAAGAGCTGA 66

RESULT 10
US-10-260-238-5427
Sequence 5427, Application US/10260238
GENERAL INFORMATION:
APPLICANT: Budworth, Paul R.
APPLICANT: Moughamer, Todd G.
APPLICANT: Briggs, Steven P.
APPLICANT: Cooper, Bret
APPLICANT: Glazedbrook, Jane
APPLICANT: Goff, Stephen A.
APPLICANT: Katagiri, Fumiyaki
APPLICANT: Kreps, Joel
APPLICANT: Provart, Nicholas
APPLICANT: Rieke, Darrell

APPLICANT: Zhu, Tong
TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
FILE REFERENCE: 60111-NP
CURRENT APPLICATION NUMBER: US/10/260,238
CURRENT FILING DATE: 2002-09-26
PRIOR APPLICATION NUMBER: US 60/325,448
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: US 60/325,277
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: US 60/370,620
PRIOR FILING DATE: 2002-04-04
NUMBER OF SEQ ID NOS: 6077
SEQ ID NO 5427
LENGTH: 546
TYPE: DNA
ORGANISM: Zea mays
US-10-260-238-5427

Query Match 46.9%; Score 28.6; DB 7; Length 546;
Best Local Similarity 72.5%; Pred. No. 22;
Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Oy 11 GAAGAACTGCTCAAGAAACCGCTGCTCAAGAACTGCTCAAGAACTGTA 61
Db 205 GATGAACTGTTGAAGAGCAGCAGCTCGAAGAACTATAGAAGAACTGGA 255

RESULT 11
US-10-260-238-4613
Sequence 4613, Application US/10260238
Publication No. US20040016025A1
GENERAL INFORMATION:

APPLICANT: Budworth, Paul R.
APPLICANT: Moughamer, Todd G.
APPLICANT: Briggs, Steven P.
APPLICANT: Cooper, Bret
APPLICANT: Glazebrook, Jane
APPLICANT: Goff, Stephen A.
APPLICANT: Katagiri, Fumiyaki
APPLICANT: Kreps, Joel
APPLICANT: Provart, Nicholas
APPLICANT: Rieke, Darrell
APPLICANT: Zhu, Tong
TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
FILE REFERENCE: 60111-NP
CURRENT APPLICATION NUMBER: US/10/260,238
CURRENT FILING DATE: 2002-09-26
PRIOR APPLICATION NUMBER: US 60/325,448
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: US 60/325,277
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: US 60/370,620
PRIOR FILING DATE: 2002-04-04
NUMBER OF SEQ ID NOS: 6077
SEQ ID NO 4613
LENGTH: 785
TYPE: DNA
ORGANISM: Trifolium aestivum
US-10-260-238-4613

Query Match 46.9%; Score 28.6; DB 7; Length 785;
Best Local Similarity 72.5%; Pred. No. 24;
Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Oy 11 GAAGAACTGCTCAAGAAACCGCTGCTCAAGAACTGCTCAAGAACTGTA 61
Db 246 GATGAAACGCTTGAACAAGCAGCTGCTCGAAGAACTATAGAAGAACTGGA 296

RESULT 12
US-10-425-114-21330
Sequence 21330, Application US/10425114
Publication No. US20040034888A1

GENERAL INFORMATION:
APPLICANT: Liu, Jingdong
APPLICANT: Zhou, Yihua
APPLICANT: Kovalic, David K.
APPLICANT: Screen, Steven E
APPLICANT: Tabaska, Jack E
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
FILE REFERENCE: 38-21(5313)B
CURRENT APPLICATION NUMBER: US/10/425,114
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 73128
SEQ ID NO 21330
LENGTH: 866
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: LIB3279-232-C9_FLI
US-10-425-114-21330

Query Match 46.9%; Score 28.6; DB 7; Length 866;
Best Local Similarity 72.5%; Pred. No. 25;
Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Oy 11 GAAGAACTGCTCAAGAAACCGCTGCTCAAGAACTGCTCAAGAACTGTA 61
Db 346 GATGAACTGTTGAAGAGCAGCAGCTCGAAGAACTATAGAAGAACTGGA 396

RESULT 13
US-10-260-238-4614
Sequence 4614, Application US/10260238
Publication No. US20040016025A1
GENERAL INFORMATION:

APPLICANT: Budworth, Paul R.
APPLICANT: Moughamer, Todd G.
APPLICANT: Briggs, Steven P.
APPLICANT: Cooper, Bret
APPLICANT: Glazebrook, Jane
APPLICANT: Goff, Stephen A.
APPLICANT: Katagiri, Fumiyaki
APPLICANT: Kreps, Joel
APPLICANT: Provart, Nicholas
APPLICANT: Rieke, Darrell
APPLICANT: Zhu, Tong
TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
FILE REFERENCE: 60111-NP
CURRENT APPLICATION NUMBER: US/10/260,238
CURRENT FILING DATE: 2002-09-26
PRIOR APPLICATION NUMBER: US 60/325,448
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: US 60/325,277
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: US 60/370,620
PRIOR FILING DATE: 2002-04-04
NUMBER OF SEQ ID NOS: 6077
SEQ ID NO 4614
LENGTH: 1019
TYPE: DNA
ORGANISM: Trifolium aestivum
FEATURE:
NAME/KEY: N region
LOCATION: (31)..(31)
OTHER INFORMATION: n = any nucleotide
FEATURE:
NAME/KEY: N region
LOCATION: (117)..(136)
OTHER INFORMATION: n = any nucleotide
US-10-260-238-4614

Query Match 46.9%; Score 28.6; DB 7; Length 1019;
Best Local Similarity 72.5%; Pred. No. 26;

Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 11 GAAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAGAACTGTA 61
DB 349 GATGAACAGCTTGAAACAAGCAGCTGCTCGAAGAGCTATGAGAGAGCTGGA 399

RESULT 14

US-10-425-115-59277
; Sequence 59277, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 59277
; LENGTH: 1166
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_154057C.1
US-10-425-115-59277

Query Match 46.9%; Score 28.6; DB 8; Length 1166;
Best Local Similarity 72.5%; Pred. No. 26;
Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 11 GAAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAGAACTGTA 61
DB 428 GATGAACAGCTTGAAAGAGCAGCTCGAAGAGCTATGAGAGAGCTGGA 478

RESULT 15

US-10-425-115-59276
; Sequence 59276, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 59276
; LENGTH: 1197
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1197)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_154056C.1
US-10-425-115-59276

Query Match 46.9%; Score 28.6; DB 8; Length 1197;
Best Local Similarity 72.5%; Pred. No. 27;
Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 11 GAAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAGAACTGTA 61
DB 474 GATGAACAGCTTGAAAGAGCAGCTCGAAGAGCTATGAGAGAGCTGGA 524

Search completed: February 27, 2006, 07:58:44
Job time: 67.7965 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:37:00 ; Search time 36.7602 Seconds
(without alignments) 3538.457 Million cell updates/sec

Title: US-10-789-164-4

Perfect score: 61
Sequence: 1 gatccaagtgcgaagaactg.....aagctgcctcaagaactgta 61

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 7209121 seqs, 1066183437 residues

Total number of hits satisfying chosen parameters: 14418242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

Published Applications NA New:*

- 1: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	27.8	45.6	588	US-09-925-065A-246263
2	27.4	44.9	2654	US-10-505-263-1
3	27.4	44.9	2748	US-10-505-263-3
4	26.4	43.3	360	US-09-925-065A-566617
5	26.4	43.3	360	US-09-925-065A-566618
6	26.4	43.3	360	US-09-925-065A-566619
7	25.8	42.3	575	US-09-925-065A-318258
8	25.6	42.0	793	US-10-750-185-55250
9	25.6	42.0	793	US-10-750-623-55250
10	25.4	41.6	3388	US-11-077-550-41
11	25.4	41.6	3388	US-11-077-550-47
12	25.4	41.6	3388	US-11-077-550-51
13	25.4	41.6	3388	US-11-077-550-55
14	25.4	41.6	3397	US-11-077-550-45
15	25.4	41.6	5292	US-10-775-169-330
16	25.2	41.3	1175	US-11-102-026A-111
17	25.2	41.3	1215	US-11-102-026A-187
18	25.2	41.3	108214	US-11-117-187-211
19	25	41.0	1008	US-11-196-475-130
20	25	41.0	1008	US-11-196-475-131

21	25	41.0	1011	US-11-196-475-51	Sequence 51, Appl
22	25	41.0	2460	US-11-072-512-37	Sequence 37, Appl
23	24.8	40.7	522	US-09-925-065A-367960	Sequence 367960, A
24	24.8	40.7	1226	US-10-750-185-54946	Sequence 54946, A
25	24.8	40.7	1226	US-10-750-623-54946	Sequence 54946, A
26	24.8	40.7	1240	US-09-925-065A-553111	Sequence 553111, A
27	24.6	40.3	390	US-09-925-065A-154591	Sequence 154591, A
28	24.6	40.3	2430	US-10-649-457-3	Sequence 3, Appl
29	24.4	40.0	200	US-11-098-666-6643	Sequence 6643, Ap
30	24.4	40.0	603	US-09-925-065A-500955	Sequence 500955, A
31	24.4	40.0	603	US-09-925-065A-500956	Sequence 500956, A
32	24.4	40.0	677	US-09-925-065A-92084	Sequence 92084, Ap
33	24.4	40.0	1251	US-11-098-686-9738	Sequence 9738, Ap
34	24.4	40.0	1457619	US-11-098-686-8739	Sequence 8739, Ap
35	24.2	39.7	609	US-10-750-623-46785	Sequence 46785, A
36	24.2	39.7	609	US-10-750-623-46785	Sequence 46785, A
37	24.2	39.7	643	US-09-925-065A-801762	Sequence 801762, A
38	24.2	39.7	643	US-09-925-065A-801763	Sequence 801763, A
39	24	39.3	1008	US-11-196-475-53	Sequence 12, Appl
40	24	39.3	1773	US-11-040-218-12	Sequence 77, Appl
41	24	39.3	2557	US-11-121-086-49	Sequence 49, Appl
42	23.8	39.0	507	US-09-925-065A-228106	Sequence 228106, A
43	23.8	39.0	577	US-09-925-065A-888138	Sequence 888138, A
44	23.8	39.0	577	US-09-925-065A-888138	Sequence 888138, A
45	23.8	39.0	589	US-09-925-065A-160795	Sequence 160795, A

ALIGNMENTS

RESULT 1
US-09-925-065A-246263
Sequence 246263, Application US/09925065A
Publication No. US20040181048A1
GENERAL INFORMATION:
APPLICANT: Manq, David G.
TITLE OF INVENTION: Identification and Mapping of Single
Nucleotide Polymorphisms in the Human Genome
FILE REFERENCE: 108827.135
CURRENT APPLICATION NUMBER: US/09/925, 065A
PRIOR FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: US 60/243, 096
PRIOR FILING DATE: 2000-10-24
PRIOR FILING DATE: 2000-11-20
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/250, 092
PRIOR FILING DATE: 2000-11-30
PRIOR APPLICATION NUMBER: US 60/261, 766
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/289, 846
NUMBER OF SEQ ID NOS: 957086
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 246263
LENGTH: 588
TYPE: DNA
ORGANISM: Homo sapiens
US-09-925-065A-246263
Query Match 45.6%; Score 27.8; DB 6; Length 588;
Best Local Similarity 74.5%; Pred. No. 2.5;
Matches 35; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
Qy 6 AAGTGAAGAACTGCTCAAGAAACCGTCTCAAGAACTGCTCA 52
Db 305 ATGTGTAAGAGGTGCCCAAGAACTGCTGTAATTAATGAA 351
RESULT 2
US-10-505-263-1
Sequence 1, Application US/10505263
Publication No. US20060014940A1
GENERAL INFORMATION:

```

; APPLICANT: Vanderbilt University
; APPLICANT: Case Western Reserve Hospital, Inc.
; APPLICANT: The Brigham and Women's Hospital, Inc.
; APPLICANT: Mount, David B
; APPLICANT: Romero, Michael
; TITLE OF INVENTION: CLONING AND CHARACTERIZATION OF SLC26A6, SLC26A1, and SLC26A2
; FILE REFERENCE: 1242/50/2 PCT/US
; CURRENT APPLICATION NUMBER: US/10/505,263
; CURRENT FILING DATE: 2004-08-20
; PRIOR APPLICATION NUMBER: US 60/360,275
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: PCT/US03/06469
; PRIOR FILING DATE: 2003-02-28
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 2654
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-505-263-1
```

```

Query Match          44.9%; Score 27.4; DB 7; Length 2654;
Best Local Similarity 75.6%; Pred. No. 5;
Matches 34; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
```

```
Qy      3  TCCAGTGAAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTG 47
Db      1778 TCCGAGAGAAAGAACTGCTCAAGAGAGGAGGACGCTGAAAGCTG 1822
```

```

RESULT 3
US-10-505-263-3
; Sequence 3, Application US/10505263
; Publication No. US20060014940A1
; GENERAL INFORMATION:
; APPLICANT: Vanderbilt University
; APPLICANT: Case Western Reserve Hospital, Inc.
; APPLICANT: The Brigham and Women's Hospital, Inc.
; APPLICANT: Mount, David B
; APPLICANT: Romero, Michael
; TITLE OF INVENTION: CLONING AND CHARACTERIZATION OF SLC26A6, SLC26A1, and SLC26A2
; FILE REFERENCE: 1242/50/2 PCT/US
; CURRENT APPLICATION NUMBER: US/10/505,263
; CURRENT FILING DATE: 2004-08-20
; PRIOR APPLICATION NUMBER: US 60/360,275
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: PCT/US03/06469
; PRIOR FILING DATE: 2003-02-28
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 2748
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (252)..(2468)
US-10-505-263-3
```

```

Query Match          44.9%; Score 27.4; DB 7; Length 2748;
Best Local Similarity 75.6%; Pred. No. 5.1;
Matches 34; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
```

```
Qy      3  TCCAGTGAAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTG 47
Db      1908 TCCGAGAGAAAGAACTGCTCAAGAGAGGAGGACGCTGAAAGCTG 1952
```

```

RESULT 4
US-09-925-065A-566617
; Sequence 566617, Application US/09925065A
```

```

; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 566617
; LENGTH: 360
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-566617
```

```

Query Match          43.3%; Score 26.4; DB 6; Length 360;
Best Local Similarity 69.2%; Pred. No. 6.9;
Matches 36; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
```

```
Qy      10  GGAAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAGAACTGTA 61
Db      140 GGAGGAGCTGTTCACAAAACAGCTTTCAGGACGCTGCTGGAATCTGTA 191
```

```

RESULT 5
US-09-925-065A-566618
; Sequence 566618, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 566618
; LENGTH: 360
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-566618
```

```

Query Match          43.3%; Score 26.4; DB 6; Length 360;
Best Local Similarity 69.2%; Pred. No. 6.9;
Matches 36; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
```

```
Qy      10  GGAAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAGAACTGTA 61
Db      140 GGAGGAGCTGTTCACAAAACAGCTTTCAGGACGCTGCTGGAATCTGTA 191
```

RESULT 6

US-09-925-065A-566619
; Sequence 566619, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 566619
; LENGTH: 360
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-566619

Query Match 43.3%; Score 26.4; DB 6; Length 360;
Best Local Similarity 69.2%; Pred. No. 6.9; Indels 0; Gaps 0;
Matches 36; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

10 GGAGAGAACTGCTCAGAAACCGCTGCTCAGAGAGCTGCTCAGAAAGCTGTA 61
DB 140 GGAGAGACTGTTCACAAAACAGCTTCCAGGACCTGCTGGAATCTGTA 191

RESULT 7

US-09-925-065A-318258
; Sequence 318258, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 318258
; LENGTH: 575
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-318258

Query Match 42.3%; Score 25.8; DB 6; Length 575;
Best Local Similarity 67.9%; Pred. No. 13;
Matches 36; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

6 AAGTGAAGAACTGCTCAGAAACCGCTGCTCAGAGAGCTGCTCAGAAAGCT 58

DB 232 AAGTGTGTGAACACCTGAGAAAAAGTTGTGAAAAAGACGCTGAGAAAGT 284

RESULT 8

US-10-750-185-55250/c
; Sequence 55250, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 55250
; LENGTH: 793
; TYPE: DNA
; ORGANISM: Bovine 19866881164033
US-10-750-185-55250

Query Match 42.0%; Score 25.6; DB 8; Length 793;
Best Local Similarity 70.8%; Pred. No. 16;
Matches 34; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

2 ATCCAAGTGAAGAAAGCTGCTCAGAAACCGCTGCTCAGAAAGCTGCT 49
DB 142 ATCCAAGCTGAGAAACATGCTGAGAACTGCTGCAAGTGAAGCTTGT 95

RESULT 9

US-10-750-623-55250/c
; Sequence 55250, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 55250
; LENGTH: 793
; TYPE: DNA
; ORGANISM: Bovine 19866881164033
US-10-750-623-55250

Query Match 42.0%; Score 25.6; DB 8; Length 793;
Best Local Similarity 70.8%; Pred. No. 16;
Matches 34; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

2 ATCCAAGTGAAGAAAGCTGCTCAGAAACCGCTGCTCAGAAAGCTGCT 49
DB 142 ATCCAAGCTGAGAAACATGCTGAGAACTGCTGCAAGTGAAGCTTGT 95

```
RESULT 10
US-11-077-550-41
; Sequence 41, Application US/11077550
; Publication No. US20050244435A1
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford Charles
; APPLICANT: Quinn, Conrad Padraig
; APPLICANT: Foster, Keith Alan
; APPLICANT: Chaddock, John
; APPLICANT: Marks, Philip
; APPLICANT: Sutton, J. Mark
; APPLICANT: Stancombe, Patrick
; APPLICANT: Wayne, Jonathan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; FILE REFERENCE: 1581.0130004
; CURRENT APPLICATION NUMBER: US/11/077,550
; PRIOR FILING DATE: 2005-03-11
; PRIOR APPLICATION NUMBER: 10/241,596
; PRIOR FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: 09/255,829
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: PCT/GB97/02273
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 08/782,893
; PRIOR FILING DATE: 1996-12-27
; PRIOR APPLICATION NUMBER: GB9625996.5
; PRIOR FILING DATE: 1996-12-27
; PRIOR APPLICATION NUMBER: GB9617671.4
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 41
; LENGTH: 3388
; TYPE: DNA
; ORGANISM: Clostridium botulinum
US-11-077-550-41

Query Match      41.6%; Score 25.4; DB 11; Length 3388;
Best Local Similarity 64.4%; Pred. No. 27;
Matches 38; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Cy      2 ATCCAAGTGAAGAAACGCTGCTCAAGAAACGCTGCTCAAGAAACTGT 60
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      2608 ATCAAGCTGAAGCGCGCGCTAAAGAGCGCGCTAAAGAGCGCGCTAAAGCGGT 2666

RESULT 11
US-11-077-550-47
; Sequence 47, Application US/11077550
; Publication No. US20050244435A1
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford Charles
; APPLICANT: Quinn, Conrad Padraig
; APPLICANT: Foster, Keith Alan
; APPLICANT: Chaddock, John
; APPLICANT: Marks, Philip
; APPLICANT: Sutton, J. Mark
; APPLICANT: Stancombe, Patrick
; APPLICANT: Wayne, Jonathan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; FILE REFERENCE: 1581.0130004
; CURRENT APPLICATION NUMBER: US/11/077,550
; PRIOR FILING DATE: 2005-03-11
; PRIOR APPLICATION NUMBER: 10/241,596
; PRIOR FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: 09/255,829
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: PCT/GB97/02273
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 08/782,893
; PRIOR FILING DATE: 1996-12-27
; PRIOR APPLICATION NUMBER: GB9625996.5
; PRIOR FILING DATE: 1996-12-13
```

```
; PRIOR APPLICATION NUMBER: GB9617671.4
; PRIOR FILING DATE: 1996-08-23
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 47
; LENGTH: 3388
; TYPE: DNA
; ORGANISM: Clostridium botulinum
US-11-077-550-47

Query Match      41.6%; Score 25.4; DB 11; Length 3388;
Best Local Similarity 64.4%; Pred. No. 27;
Matches 38; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Cy      2 ATCCAAGTGAAGAAACGCTGCTCAAGAAACGCTGCTCAAGAAACTGT 60
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      2608 ATCAAGCTGAAGCGCGCGCTAAAGAGCGCGCGCTAAAGAGCGCGCTAAAGCGGT 2666

RESULT 12
US-11-077-550-51
; Sequence 51, Application US/11077550
; Publication No. US20050244435A1
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford Charles
; APPLICANT: Quinn, Conrad Padraig
; APPLICANT: Foster, Keith Alan
; APPLICANT: Chaddock, John
; APPLICANT: Marks, Philip
; APPLICANT: Sutton, J. Mark
; APPLICANT: Stancombe, Patrick
; APPLICANT: Wayne, Jonathan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; FILE REFERENCE: 1581.0130004
; CURRENT APPLICATION NUMBER: US/11/077,550
; PRIOR FILING DATE: 2005-03-11
; PRIOR APPLICATION NUMBER: 10/241,596
; PRIOR FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: 09/255,829
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: PCT/GB97/02273
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: 08/782,893
; PRIOR FILING DATE: 1996-12-27
; PRIOR APPLICATION NUMBER: GB9625996.5
; PRIOR FILING DATE: 1996-12-13
; PRIOR APPLICATION NUMBER: GB9617671.4
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 51
; LENGTH: 3388
; TYPE: DNA
; ORGANISM: Clostridium botulinum
US-11-077-550-51

Query Match      41.6%; Score 25.4; DB 11; Length 3388;
Best Local Similarity 64.4%; Pred. No. 27;
Matches 38; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Cy      2 ATCCAAGTGAAGAAACGCTGCTCAAGAAACGCTGCTCAAGAAACTGT 60
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      2608 ATCAAGCTGAAGCGCGCGCTAAAGAGCGCGCGCTAAAGAGCGCGCTAAAGCGGT 2666

RESULT 13
US-11-077-550-55
; Sequence 55, Application US/11077550
; Publication No. US20050244435A1
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford Charles
; APPLICANT: Quinn, Conrad Padraig
; APPLICANT: Foster, Keith Alan
```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:34:23 ; Search time 14.3542 Seconds
(without alignments)
7553.984 Million cell updates/sec

Title: US-10-789-164-4

Perfect score: 61
Sequence: 1 gatccaagtcgaagaactg.....aagctcctcaagaactgta 61

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

Issued Patents NA: *
1: /cgn2_6/prodata/1/ina/1/COMB.seq:*
2: /cgn2_6/prodata/1/ina/5/COMB.seq:*
3: /cgn2_6/prodata/1/ina/6/COMB.seq:*
4: /cgn2_6/prodata/1/ina/8/COMB.seq:*
5: /cgn2_6/prodata/1/ina/H/COMB.seq:*
6: /cgn2_6/prodata/1/ina/CTTS/COMB.seq:*
7: /cgn2_6/prodata/1/ina/PP/COMB.seq:*
8: /cgn2_6/prodata/1/ina/RP/COMB.seq:*
9: /cgn2_6/prodata/1/ina/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	30.4	49.8	882 3 US-10-166-653-11	Sequence 11, Appl
2	28.4	46.6	354 3 US-09-248-796A-5595	Sequence 5595, Ap
3	28.2	46.2	1920 2 US-08-186-222-1	Sequence 1, Appl1
4	27.8	45.6	2103 3 US-09-107-532A-1440	Sequence 1440, Ap
5	27.4	44.6	1466 3 US-09-620-312D-914	Sequence 914, Appl
6	27.2	44.6	47115 3 US-09-949-016-12278	Sequence 12278, A
7	27.2	44.6	47122 3 US-09-949-016-16520	Sequence 16520, A
8	26.8	43.9	5962 9 5386025-5	Patent No. 5386025
9	26.8	43.9	5975 2 US-08-404-354B-1	Sequence 1, Appl1
10	26.8	43.9	5975 2 US-08-314-083B-1	Sequence 1, Appl1
11	26.8	43.9	5975 2 US-08-435-675B-1	Sequence 1, Appl1
12	26.8	43.9	5975 2 US-08-336-257A-3	Sequence 3, Appl1
13	26.8	43.9	5975 2 US-08-884-599-1	Sequence 1, Appl1
14	26.6	43.6	11440 3 US-09-328-352-1093	Sequence 1093, Ap
15	26.4	43.3	1324 3 US-09-902-540-7007	Sequence 1007, Ap
16	26.4	43.3	5164 3 US-09-902-540-615	Sequence 615, Appl
17	26.2	43.0	516 3 US-09-270-767-753B	Sequence 753B, Ap
18	26.2	43.0	516 3 US-09-270-767-753B	Sequence 22820, A
19	26.2	43.0	723 3 US-09-603-208A-9	Sequence 9, Appl1
20	26.2	43.0	1548 3 US-09-248-796A-6167	Sequence 6167, Ap
21	25.6	42.0	528 3 US-09-248-796A-4819	Sequence 4819, Ap
22	25.6	42.0	1131 3 US-09-248-796A-994	Sequence 994, Appl
23	25.2	41.3	345 3 US-09-134-000C-1925	Sequence 1925, Ap
24	25.2	41.3	1658 3 US-09-016-434-1320	Sequence 1320, Ap

25	25.2	41.3	2843 3 US-09-620-312D-211	Sequence 211, Appl
26	25.2	41.3	3218 3 US-09-799-451-470	Sequence 470, Appl
27	25.2	41.3	7176 3 US-09-221-017B-626	Sequence 626, Appl
28	25.2	41.3	1230025 3 US-09-198-452A-1	Sequence 1, Appl1
29	25.2	41.3	1230230 3 US-09-438-185A-1	Sequence 1092, Appl
30	25	41.0	912 3 US-09-540-236-1092	Sequence 130, Appl
31	25	41.0	1008 3 US-08-235-836C-130	Sequence 131, Appl
32	25	41.0	1008 3 US-08-235-836C-131	Sequence 51, Appl
33	25	41.0	1011 3 US-08-235-836C-51	Sequence 625, Appl
34	25	41.0	1548 3 US-09-799-451-625	Sequence 5, Appl1
35	25	41.0	1945 2 US-08-525-742-5	Sequence 12709, A
36	25	41.0	2430 3 US-09-252-991A-12709	Sequence 37, Appl
37	25	41.0	2460 3 US-10-104-047-37	Sequence 13249, A
38	25	41.0	2607 3 US-09-252-991A-13249	Sequence 139, Appl
39	24.8	40.7	501 3 US-09-270-767-15621	Sequence 15621, A
40	24.8	40.7	501 3 US-09-270-767-15621	Sequence 19228, A
41	24.6	40.3	200 3 US-09-513-999C-19228	Sequence 3082, Appl
42	24.6	40.3	375 3 US-09-134-000C-3082	Sequence 3, Appl1
43	24.6	40.3	1209 3 US-09-464-483-3	Sequence 3, Appl1
44	24.6	40.3	1209 3 US-09-414-664-3	Sequence 1, Appl1
45	24.6	40.3	1589 3 US-09-464-483-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1
US-10-166-653-11
Sequence 11, Application US/1016653
Patent No. 6911332
GENERAL INFORMATION:
APPLICANT: USUDA, Yoshihiro
APPLICANT: NISHIO, Yosuke
APPLICANT: YASUEDA, Hisashi
APPLICANT: SUGIMOTO, Shinichi
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING POLYPEPTIDES INVOLVED IN ONE-CARBON COMPOUND METABOLISM IN METHYLOPHILUS METHYLOTROPHUS
FILE REFERENCE: 211926USO
CURRENT APPLICATION NUMBER: US/10166, 653
CURRENT FILING DATE: 2002-06-12
NUMBER OF SEQ ID NOS: 40
SOFTWARE: Patentin version 3.1
SEQ ID NO 11
LENGTH: 882
TYPE: DNA
ORGANISM: Methylophilus methylotrophus
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(882)
OTHER INFORMATION:
US-10-166-653-11
Query Match 49.8%; Score 30.4; DB 3; Length 882;
Best Local Similarity 71.4%; Pred. No. 0.89; 16; Indels 0; Gaps 0;
Matches 40; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
OY 6 AACTGGAAGAACTGCTCAAGAAACCGCTGCTCAAGAACTGCTCAAGAACTGTA 61
Db 815 AAGAGCTGACCTGCTGAAGCTGCGCTGCTGAAGAACTGCGCAACTGAA 870
RESULT 2
US-09-248-796A-5595
Sequence 5595, Application US/09248796A
Patent No. 6747137
GENERAL INFORMATION:
APPLICANT: Keith Weinstein et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
FILE REFERENCE: 107196.132
CURRENT APPLICATION NUMBER: US/09248, 796A
CURRENT FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 60/074, 725

PRIOR FILING DATE: 1998-02-13
PRIOR APPLICATION NUMBER: US 60/096,409
PRIOR FILING DATE: 1998-08-13
NUMBER OF SEQ ID NOS: 28208
SEQ ID NO: 5595
LENGTH: 354
TYPE: DNA
ORGANISM: *Candida albicans*
US-09-248-796A-5595

Query Match 46.6%; Score 28.4; DB 3; Length 354;
Best Local Similarity 76.1%; Pred. No. 3.3;
Matches 35; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 11 GAAGAACTGCTCAGAAACCGCTGCTCAGAGAGCTGCTCAGAA 56
DB 286 GAAGAACTGCTCAGAAACCGCTGCTCAGAGAGCTGCTCAGAA 331

RESULT 3

US-08-186-222-1
Sequence 1, Application US/08186222
Patent No. 5559007

GENERAL INFORMATION:

APPLICANT: Surti, Bruno
APPLICANT: Schmitz, Albert
TITLE OF INVENTION: Bacterial Vectors
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/186,222
FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/672,205
FILING DATE: 19-MAR-1991
APPLICATION NUMBER: GB 9006400.7
FILING DATE: 22-MAR-1990

ATTORNEY/AGENT INFORMATION:
NAME: Villamizar, Joann
REGISTRATION NUMBER: 30,598
REFERENCE/DOCKET NUMBER: 4-17994/A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914)785-7121
TELEFAX: (914)347-5769

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 1920 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

ORIGINAL SOURCE:

ORGANISM: *Lactococcus lactis* LMO230
INDIVIDUAL ISOLATE: Major Secretion Product (MSP) Gene
IMMEDIATE SOURCE:
CLONE: pUCS

FEATURE:

NAME/KEY: CDS
LOCATION: 411..1793

US-08-186-222-1

Query Match 46.2%; Score 28.2; DB 2; Length 1920;
Best Local Similarity 73.5%; Pred. No. 5.8;
Matches 36; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 7 AGTGAAGAACTGCTCAGAAACCGCTGCTCAGAGAGCTGCTCAGAA 55
DB 1061 AGATGAAGAAACGAGCTCAGAAACCGCTGCTCAGAGAGCTGCTCAGAA 1109

RESULT 4

US-09-107-532A-1440
Sequence 1440, Application US/09107532A
Patent No. 6583275

GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-8277
TELEFAX: (781)893-5007

INFORMATION FOR SEQ ID NO: 1440:
SEQUENCE CHARACTERISTICS:
LENGTH: 2103 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:

ORGANISM: *Enterococcus faecium*
FEATURE:

NAME/KEY: misc_feature
LOCATION: (B) LOCATION 1..2103
SEQUENCE DESCRIPTION: SEQ ID NO: 1440:

US-09-107-532A-1440

Query Match 45.6%; Score 27.8; DB 3; Length 2103;
Best Local Similarity 69.1%; Pred. No. 8;
Matches 38; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 5 CAAGTGAAGAACTGCTCAGAAACCGCTGCTCAGAGAGCTGCTCAGAACTG 59
DB 1687 CAAGCAAAACGAGCTCAGAAACCGCTGCTCAGAGAGCTGCTCAGAACTG 1741

RESULT 5

US-09-620-312D-914

```
; Sequence 914, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yuning
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhixue
; APPLICANT: John, Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; FILE REFERENCE: 784C1P2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pc_fl_genes Version 1.0
; SEQ ID NO 914
; LENGTH: 1466
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (779)..(1213)
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(1466)
; OTHER INFORMATION: n = a,t,c or g
; US-09-620-312D-914

Query Match          44.9%; Score 27.4; DB 3; Length 1466;
Best Local Similarity 75.6%; Pred. No. 9;
Matches 34; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy          3 TCCAACTGGAAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTG 47
Db          418 TCCCAAGAGAAAGAACTGCTCAAGAAAGCAGGACGCTGAAGCTG 462

RESULT 6
US-09-949-016-12278/c
; Sequence 12278, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12278
```

```
; LENGTH: 47115
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(47115)
; OTHER INFORMATION: n = A,T,C or G
; US-09-949-016-12278

Query Match          44.6%; Score 27.2; DB 3; Length 47115;
Best Local Similarity 67.9%; Pred. No. 28;
Matches 38; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy          1 GATCCAAGTGAAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAGAA 56
Db          39716 GATCAAAATGAATCAAACTGGCCAGGCAATGCTCTCAACAGATGCTCAACAA 39661

RESULT 7
US-09-949-016-16520/c
; Sequence 16520, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16520
; LENGTH: 47122
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(47122)
; OTHER INFORMATION: n = A,T,C or G
; US-09-949-016-16520

Query Match          44.6%; Score 27.2; DB 3; Length 47122;
Best Local Similarity 67.9%; Pred. No. 28;
Matches 38; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy          1 GATCCAAGTGAAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAGAA 56
Db          39716 GATCAAAATGAATCAAACTGGCCAGGCAATGCTCTCAACAGATGCTCAACAA 39661

RESULT 8
US-09-949-016-16520/c
; Sequence 16520, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16520
; LENGTH: 47122
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(47122)
; OTHER INFORMATION: n = A,T,C or G
; US-09-949-016-16520

Query Match          44.6%; Score 27.2; DB 3; Length 47122;
Best Local Similarity 67.9%; Pred. No. 28;
Matches 38; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy          1 GATCCAAGTGAAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAGAA 56
Db          39716 GATCAAAATGAATCAAACTGGCCAGGCAATGCTCTCAACAGATGCTCAACAA 39661

RESULT 5
US-09-949-016-16520/c
; Sequence 16520, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16520
; LENGTH: 47122
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(47122)
; OTHER INFORMATION: n = A,T,C or G
; US-09-949-016-16520

Query Match          43.9%; Score 26.8; DB 9; Length 5962;
Best Local Similarity 73.9%; Pred. No. 22;
Matches 34; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
```

QY 8 GTGGAAGAACTGCTCAAGAAACGCTGCTCAAGAGCTGCTCAAG 53
DB 111 GAGGAGAAACAGCCCAAGAAAGCCCTGCTCCGAGGTCTGCTCCAGG 156

RESULT 9

US-08-404-354B-1
Sequence 1, Application US/08404354B.
Patent No. 5618720
GENERAL INFORMATION:
APPLICANT: Ellis, Steven Bradley
APPLICANT: Williams, Mark E.
APPLICANT: Harpold, Michael Miller
APPLICANT: Schwartz, Robert
APPLICANT: Brenner, Robert
TITLE OF INVENTION: CALCIUM CHANNEL COMPOSITIONS AND METHODS
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Martin, Haller & McClain
STREET: 1660 Union Street
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92101-2926
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/404,354B
FILING DATE: 15-FEB-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,083
FILING DATE: 28-SEP-1994
APPLICATION NUMBER: US 07/914,231
FILING DATE: 13-JUL-1992
APPLICATION NUMBER: US 07/603,751
FILING DATE: 08-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 6362-53192
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-238-0999
TELEFAX: 619-238-0062
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5975 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 79...5700
OTHER INFORMATION:
US-08-404-354B-1

Query Match 43.9%; Score 26.8; DB 2; Length 5975;
Best Local Similarity 73.9%; Pred. No. 22;
Matches 34; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 8 GTGGAAGAACTGCTCAAGAAACGCTGCTCAAGAGCTGCTCAAG 53
DB 111 GAGGAGAAACAGCCCAAGAAAGCCCTGCTCCGAGGTCTGCTCCAGG 156

RESULT 10
US-08-314-083B-1

Sequence 1, Application US/08314083B
Patent No. 5686241
GENERAL INFORMATION:
APPLICANT: Ellis, Steven Bradley
APPLICANT: Williams, Mark E.
APPLICANT: Harpold, Michael Miller
APPLICANT: Schwartz, Robert
APPLICANT: Brenner, Robert
TITLE OF INVENTION: CALCIUM CHANNEL COMPOSITIONS AND METHODS
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Martin, Haller & McClain
STREET: 1660 Union Street
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92101-2926
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/314,083B
FILING DATE: 28-SEP-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/914,231
FILING DATE: 13-JUL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/603,751
FILING DATE: 08-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 6362-53191
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-238-0999
TELEFAX: 619-238-0062
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5975 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 79...5700
OTHER INFORMATION:
US-08-314-083B-1

Query Match 43.9%; Score 26.8; DB 2; Length 5975;
Best Local Similarity 73.9%; Pred. No. 22;
Matches 34; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 8 GTGGAAGAACTGCTCAAGAAACGCTGCTCAAGAGCTGCTCAAG 53
DB 111 GAGGAGAAACAGCCCAAGAAAGCCCTGCTCCGAGGTCTGCTCCAGG 156

RESULT 11
US-08-435-675B-1
Sequence 1, Application US/08435675B
Patent No. 5710250
GENERAL INFORMATION:
APPLICANT: Ellis, Steven Bradley
APPLICANT: Williams, Mark E.
APPLICANT: Harpold, Michael Miller
APPLICANT: Schwartz, Robert
APPLICANT: Brenner, Robert
TITLE OF INVENTION: CALCIUM CHANNEL COMPOSITIONS AND METHODS

NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESS: Brown, Martin, Haller & McClain
STREET: 1660 Union Street
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92101-2926
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/435, 675B
FILING DATE: 05-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,083
FILING DATE: 28-SEP-1994
APPLICATION NUMBER: US 07/914,231
FILING DATE: 13-JUL-1992
APPLICATION NUMBER: US 07/603,751
FILING DATE: 08-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L.
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 6362-53193
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-238-0999
TELEFAX: 619-238-0062
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5975 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 79...5700
OTHER INFORMATION:
US-08-435-675B-1
Query Match 43.9%; Score 26.8; DB 2; Length 5975;
Best Local Similarity 73.9%; Pred. No. 22;
Matches 34; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
QY 8 GTGGAAGAACTGCTCAGAAACCGCTGCTCAAGAGCTGCTCAG 53
DB 111 GAGGAAGAAACAGCCCAAGAAAGCCCTGCGCCGAGTCTCCAGG 156
RESULT 12
US-08-336-257A-3
Sequence 3, Application US/08336257A
Patent No. 5726035
GENERAL INFORMATION:
APPLICANT: Jay, Scott D.
APPLICANT: Ellis, Steven B.
APPLICANT: Harpold, Michael M.
APPLICANT: Campbell, Kevin P.
TITLE OF INVENTION: CALCIUM CHANNEL COMPOSITIONS AND METHODS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Martin, Haller & McClain
STREET: 1660 Union Street
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92101-2926
COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/336, 257A
FILING DATE: 07-NOV-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L.
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 54898
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 238-0999
TELEFAX: (619) 238-0062
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 5975 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: Genomic DNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 79...5700
OTHER INFORMATION: "product= "Alpha-1 subunit of animal calcium channel"
US-08-336-257A-3
Query Match 43.9%; Score 26.8; DB 2; Length 5975;
Best Local Similarity 73.9%; Pred. No. 22;
Matches 34; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
QY 8 GTGGAAGAACTGCTCAGAAACCGCTGCTCAAGAGCTGCTCAG 53
DB 111 GAGGAAGAAACAGCCCAAGAAAGCCCTGCGCCGAGTCTCCAGG 156
RESULT 13
US-08-884-599-1
Sequence 1, Application US/08884599
Patent No. 6013474
GENERAL INFORMATION:
APPLICANT: Williams, Steven Bradley
APPLICANT: Williams, Mark B.
APPLICANT: Harpold, Michael Miller
APPLICANT: Schwartz, Arnold
APPLICANT: Brenner, Robert
TITLE OF INVENTION: CALCIUM CHANNEL COMPOSITIONS AND METHODS
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Martin, Haller & McClain
STREET: 1660 Union Street
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92101-2926
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/884, 599
FILING DATE: 27-JUNE-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/314,083
FILING DATE: 28-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/914,231
FILING DATE: 13-JUL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/603,751

FILING DATE: 08-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 6362-53191B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-238-0999
TELEFAX: 619-238-0062
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5975 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 79..5700
OTHER INFORMATION:
US-08-884-599-1

Query Match 43.9%; Score 26.8; DB 3; Length 5975;
Best Local Similarity 73.9%; Pred. No. 22;
Matches 34; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

OY 8 GTGGAAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAG 53
DB 111 GAGGAAGAAACGACCCCAAGAGCCCTGCGCGAGTCTGCGCCAG 156

RESULT 14
US-09-328-352-1093
Sequence 1093, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 1093
LENGTH: 1140
TYPE: DNA
ORGANISM: Acinetobacter baumannii
US-09-328-352-1093

Query Match 43.6%; Score 26.6; DB 3; Length 1140;
Best Local Similarity 71.4%; Pred. No. 17;
Matches 35; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

OY 7 AGTGAAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAGAA 55
DB 1092 AGTACTCAACTGCTCAAGAGCGGCACTCTGACGAGCTCAATTA 1140

RESULT 15
US-09-902-540-7007
Sequence 7007, Application US/09902540
Patent No. 6833447
GENERAL INFORMATION:
APPLICANT: Goldman, Barry S.
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Wiegand, Roger C.
TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
FILE REFERENCE: 38-10(15849)B
CURRENT APPLICATION NUMBER: US/09/902,540
CURRENT FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: 60/217,883
PRIOR FILING DATE: 2000-07-10

NUMBER OF SEQ ID NOS: 16825
SEQ ID NO 7007
LENGTH: 324
TYPE: DNA
ORGANISM: Myxococcus xanthus
US-09-902-540-7007

Query Match 43.3%; Score 26.4; DB 3; Length 324;
Best Local Similarity 75.0%; Pred. No. 15;
Matches 33; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

OY 12 AAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAGAA 55
DB 52 AAGCAAGAGTCAAGAAACCGCGCTCTACAGAGTCTCTGCA 95

Search completed: February 27, 2006, 06:46:21
Job time : 16.5542 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:33:05 ; Search time 368.475 Seconds
(without alignments)
9873.061 Million cell updates/sec

Title: US-10-789-164-6

Perfect score: 64
Sequence: 1 gatccgcgaagaaggtgttc.....aaatctggaactggaagta 64

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sta:*
11: gb_sy:*
12: gb_un:*
13: gb_vl:*
14: gb_hlg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	28.8	45.0	86313	5	CR558305
C 2	28.8	45.0	213324	9	AC161816
C 3	28.6	44.7	25600	2	U80442
C 4	28.6	44.7	189293	9	AC161379
C 5	28.6	44.7	201463	14	AC154577
C 6	28.6	44.7	280474	14	AC160041
C 7	28.6	44.7	294136	14	AC006901
C 8	28.6	44.4	172866	14	AC164240
C 9	28.2	44.1	145870	14	AC149957
C 10	28.2	44.1	229461	14	AC155077
C 11	27.8	43.4	186622	8	AC125238
C 12	27.4	42.8	81826	8	HSU78045
C 13	27.4	42.8	173359	14	AP000922
C 14	27.4	42.8	180399	14	AP001144
C 15	27.2	42.5	34337	6	AX398707
C 16	27.2	42.5	61550	8	AY049008
C 17	27.2	42.5	87187	8	AL161744
C 18	27.2	42.5	110000	14	AC055726_2

19	27.2	42.5	148390	9	AC121930	AC121930 Mus muscu
C 20	27.2	42.5	160008	14	AC022823	AC022823 Homo sapi
C 21	27.2	42.5	162228	8	AL450382	AL450382 Human DNA
C 22	27.2	42.5	163231	8	AL954360	AL954360 Human DNA
C 23	27.2	42.5	166256	8	AC018922	AC018922 Homo sapi
C 24	27.2	42.5	167071	8	AC128649	AC128649 Homo sapi
C 25	27.2	42.5	187130	14	AL583858	AL583858 Homo sapi
C 26	27.2	42.5	255976	9	AC147230	AC147230 Mus muscu
C 27	27.2	42.5	274996	14	AC096298	AC096298 Rattus no
C 28	27.2	42.5	324582	14	AC133362	AC133362 Rattus no
C 29	26.8	41.9	10841	1	AE013054	AE013054 Thermococ
C 30	26.8	41.9	59818	4	AC123967	AC123967 Smittophs
C 31	26.6	41.6	196149	5	AB196462	AB196462 Oncorhyn
C 32	26.6	41.6	110000	15	AE016818_01	Continuation (2 of
C 33	26.6	41.6	169794	14	AC004688	AC004688 Plasmodiu
C 34	26.6	41.6	172883	9	AC132274	AC132274 Mus muscu
C 35	26.6	41.6	196149	14	AC004709	AC004709 Plasmodiu
C 36	26.6	41.6	252650	2	AE014847	AE014847 Plasmodiu
C 37	26.4	41.2	1209	15	AK108702	AK108702 Oryza sat
C 38	26.4	41.2	1578	6	AX834701	AX834701 Sequence
C 39	26.4	41.2	1578	8	AK097338	AK097338 Homo sapi
C 40	26.4	41.2	2434	8	AY358138	AY358138 Homo sapi
C 41	26.4	41.2	31818	2	AF125442	AF125442 Caenorhab
C 42	26.4	41.2	110000	15	AP008211_237	Continuation (238
C 43	26.4	41.2	127642	15	AC093952	AC093952 Oryza sat
C 44	26.4	41.2	137357	15	AC118289	AC118289 Oryza sat
C 45	26.2	40.9	5673	15	AY026257	AY026257 Magnapor

ALIGNMENTS

RESULT 1
CR558305/c
LOCUS
DEFINITION
Zebrafish DNA sequence from clone DKRP-87C9 in linkage group 13,
complete sequence.
ACCESSION
CR558305.6 GI:55294877
VERSION
CR558305
KEYWORDS
HTG.
SOURCE
Danio rerio (zebrafish)
ORGANISM
Danio rerio
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 86313)
Hammond,S.
Direct Submission
Submitted (03-NOV-2004) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfsh-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Nov 3, 2004 this sequence version replaced gi:54606595.
Genome Center
Center: Wellcome Trust Sanger Institute
Web site: http://www.sanger.ac.uk
Contact: zfsh-help@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC. The following abbreviations are used to associate primary accession

Query Match Similarity 45.0%; Score 28.8; DB 5; Length 86313;
Best Local Similarity 82.5%; Pred. No. 57;
Matches 33; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

LOCUS	AC161816	213324 bp	DNA	linear	ROD 03-AUG-2005
DEFINITION	Mus musculus BAC clone RP23-308L3 from chromosome 16, complete sequence.				

ACCESSION	AC161816
VERSION	AC161816.5
KEYWORDS	GI:71725592
SOURCE	HTG.
ORGANISM	Mus musculus (house mouse)
	Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.

AUTHORS	Tomlinson, C., Lewis, S. and Levy, A.
TITLE	The sequence of Mus musculus BAC clone RP23-308L3
JOURNAL	Unpublished (2001)

TITLE Direct Submission

JOURNAL Submitted (20-MAY-2005) Genome Se

REFERENCE 3 (bases 1 to 213324)

AUTHORS
Wilson, R. K.

JOURNAL
Submitted (23-JUN-2005) Genome Se

REFERENCE 4 (bases 1 to 213324)

AUTHORS
WILSON, R. K.
New York, N. Y.

JOURNAL
Submitted (03-AUG-2005) Genome Se
Virus: Hepatitis B Virus (HBV)

COMMENT On Aug 3, 2005 this sequence vers

Center: Washington University Genome Sequencing Center
Genome Center
Center code: WUGSC
Web site: <http://genome.wustl.edu>
Contact: submissions@wustl.wustl.edu
----- Summary Statistics

NOTICE

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e. phred quality >=30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone, fosmid clone or direct clone walk sequence. Sequence from the Mouse Genome Sequencing Consortium whole genome shotgun may have been used to obtain the consensus sequence. The assembly was confirmed by restriction digest.

This finishing standard has slightly changed from the previous Human standard. Specifically, standards for regions of low sequence complexity (such as dinucleotide repeats and small unit tandem repeats) have been relaxed. These regions are very prevalent in the mouse genome, and the return on extended finishing efforts is minimal.

If a sequence meets the criteria of the above statement, it needs no comments or tags. If the criteria are not met, such as ambiguous bases, then the region is duly annotated.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see: <http://genome.wustl.edu>

SOURCE INFORMATION: The BAC Library has been constructed by Kazutoyo Oeegawa and Minko Tatenno in the laboratory of Pieter de Jong (<http://www.chori.org>) from female C57BL/6J mouse kidney and/or brain genomic DNA. The clone and detailed information can be obtained from Research Genetics, Inc. (<http://www.reagen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

This sequence is the entire insert of the clone

FEATURES

Source

**H
-
C
O
N
T
E
N
T**

```
/mol type="genomic DNA"
```

```
/db xref="taxon:10090"
```

```
/chromosome="16"
```

```
/clone="RP23-308L3"
```

```
clone_lib="RPCT-23"
```

144936. : 145884

/11016E...sequence derived from product of project DNA
147347 147321

```
11/23/: :11/221
/notes="Sequence derived from one plasmid subclone."
```

148786. :148957

/note="Sequence derived from one plasmid subclone."

150162. .150279

/note="Sequence derived from one plasmid subclone."

151242. .151320

```
1513357      /note="Sequence derived from one plasmid subclone."
```

161365. .161388 derived from one subclone "

172382 172405
/HOLCE= sequence delivered 2200H OHC
pharmaceuticals
SUBCATION:

/note="Sequence derived from one plasmid subclone."

174673. 174686

/note="Sequence derived from PCR product of project DNA"

175130. .175287

/note="Sequence derived from PCR product of project DNA"

206133. :206193
derived from an olivine subzone "

ORIGIN

Query Match	45.0%;	Score 28.8;	DB 9;	Length 213324;
Best Local Similarity	65.6%;	Pred. No. 59;		
Matches 42;	Conservative	0;	Mismatches 22;	Indels 0;
			Gaps	0;

Oy 1 GATCCGCAAGAGTGTTCACACCCCTGAGAGGCTGTAGCAAAATCTGGAAGCTGA 60
 Db 190239 GGTCCACAAAGAGGTGTGATGCACGAGAGGATTTCTAAAGATCAAGGTTCTGA 190298
 Oy 61 AGTA 64
 Db 190239 TCTA 190302

RESULT 3
 U80442/c 25600 bp DNA linear INV 22-SEP-2004
 LOCUS U80442
 DEFINITION Caenorhabditis elegans cosmid T20F5, complete sequence.
 ACCESSION U80442
 VERSION U80442.1 GI:1703614
 KEYWORDS HTG.
 SOURCE Caenorhabditis elegans
 ORGANISM Caenorhabditis elegans
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditoidae; Rhabditidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 25600)

REFERENCE
 AUTHORS WormBase Consortium
 CONSRTM Genome sequence of the nematode C. elegans: a platform for
 TITLE investigating biology. The C. elegans Sequencing Consortium
 JOURNAL Science 282 (5396), 2012-2018 (1998)
 PUBMED 9851916
 REFERENCE 2 (bases 1 to 25600)
 AUTHORS Du Z. and Le T.T.
 TITLE The sequence of C. elegans cosmid T20F5
 JOURNAL Unpublished (2001)
 REFERENCE 3 (bases 1 to 25600)
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (27-NOV-1996) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 4 (bases 1 to 25600)
 REFERENCE 4
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (25-JUN-2001) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 5 (bases 1 to 25600)
 REFERENCE 5
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (23-MAY-2002) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 6 (bases 1 to 25600)
 REFERENCE 6
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (06-NOV-2002) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 7 (bases 1 to 25600)
 REFERENCE 7
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (21-NOV-2002) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 8 (bases 1 to 25600)
 REFERENCE 8
 AUTHORS Wilson, R.
 TITLE Direct Submission
 JOURNAL Submitted (15-JUN-2003) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 9 (bases 1 to 25600)
 REFERENCE 9
 AUTHORS WormBase Consortium
 CONSRTM Direct Submission
 JOURNAL Submitted (22-SEP-2004) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.

COMMENT

Louis, MO 63110, USA

Submitted by:

Genome Sequencing Center
 Department of Genetics, Washington University
 St. Louis, MO 63110, USA, and
 Sanger Centre, Hinxton Hall
 Cambridge CB10 1RQ, England
 email: submissions@watson.wustl.edu and jesus@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
 It may be shorter because we only sequence overlapping sections
 once, or longer because we provide a small overlap between
 neighboring submissions.

This sequence was finished as follows unless otherwise noted: all
 regions were double stranded, sequenced with an alternate chemistry
 or covered by high quality data (i.e., phred quality >= 30); an
 attempt was made to resolve all sequencing problems, such as
 compressions and repeats; all regions were covered by sequence from
 more than one m33 subclone.

For a graphical representation of this clone sequence and its
 analysis see:
<http://www.wormbase.org/db/seq/sequence?name=T20F5;class=Sequence>

NEIGHBORING CLONE INFORMATION

The 5' clone is C50F2, 1000 bp overlap; the 3' clone is T21E3, 200
 bp overlap. Actual start of this clone is at base position 1 of
 T20F5; actual end is at 10147 of T21E3.

NOTES:

Coding sequences below are the result of integration and manual
 review of the following data: computer analysis using the program
 GeneFinder (P. Green and L. Hillier, personal communication), the
 large scale EST projects of Yujii Kohara
 (http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The C.
 elegans ORFome cloning project (<http://wormfdb.dfci.harvard.edu/>),
 similarity to other proteins from Blastx analysis
 (<http://blast.wustl.edu/>), sequence conservation with C. briggsae
 using Jim Kent's WABA alignment program (Genome Research
 10:1115-1125, 2000), individual C. elegans Genbank submissions,
 and personal communications with C. elegans Genbank researchers,
 are predicted using the program CRMapScan-SE (Lowe, T.M. and
 Eddy, S.R., 1997, Nucleic Acids. Res., 25, 955-964).

FEATURES

source

1..25600

/organism="Caenorhabditis elegans"

/mol_type="genomic DNA"

/strain="Bristol N2"

/db_xref="taxon:6239"

/chromosome="I"

/clone="T20F5"

4237..4238

/note="SL1 trans-splice acceptor; see YK34595.5"

4243..5005

/gene="pds-4"

/locus_tag="T20F5.2"

join(4243..4345,4395..4690,4741..4807,4863..5005)

/gene="pds-4"

/locus_tag="T20F5.2"

/standard_name="T20F5.2"

/note="contains similarity to Pfam domain PF00227

(Proteasome A-type and B-type); coded for by the following

C. elegans CDNAs: BE228120, CB104130, YK211c2.5,

YK335a11.5, YK345g5.3, YK345g5.5, YK416b12.3, YK416b12.5,

YK482b2.3, YK482b2.5, YK507g8.3, YK507g8.5, YK507g8.5,

YK657a8.5, YK680b1.3, YK680b1.5, YK776d09.3, YK1532c08.3,

YK1532c08.5"

/codon_start=1

/product="Proteasome beta subunit protein 4"

/protein_id="AAB37663.1"

CDS

gene

misc_feature

```

/db_xref="GI:1703615"
/db_xref=" WormBase: T20F5. 2"
/translation="MAGNHPLFISTENYVTLADKATPAVGAILDSENDXEYRLGK
KLTMWICIEGBDVAFQPMWTKRNLOLVSRNGQYFSPGCAHFPVRSIAEGLRQDHY
TVDVILIGDYDEKEDPAFGSDVYDLANGGQOPLYFGCGRFFCYAIMDEYKMDTEA
EGLMNMKICIGAEKRRFVANIIGYKVVIIDKKYRKDDVLF"
complement(5177. .7962)
/gene="T20F5.3"
/locus_tag="T20F5.3"
/locus_tag="T20F5.3"
/standard_name="T20F5.3"
/notes="contains similarity to Pfam domain PF01765
(Ribosome recycling factor), coded for by the following C.
elegans cDNAs: Yk374a2.5, Yk675a7.3, Yk675a7.5,
Yk1697e12.3"
/codon_start=1
/product="Hypothetical protein T20F5.3"
/protein_id="AAB37664.1"
/db_xref="GI:1703616"
/db_xref=" WormBase: T20F5.3"
/translation="MAESVOPTGRAVLRLELAIEIPGEPEPSSNYDAFCDRPSAE
NMETKTKYVILLILIPSLYFLAVYINPNPKRVHPPEPTVYINITPSEELSTSE
EYQAQFELAVKNDQVLPTGVSRFPMGMLRLAVIRLQAQSRAPQIQKQAQIFTS
EYNAKKRDKKKRNPAPVSNLEENAVVETIKETQREGLBELTETHFSLKDIR
YEDVAVKLENGDKPELSMARTATLQSPMLMINIPQDNPSIAKAKLAIQKTLVTPC
QEGALVYVNPSPMSKERERKMSDAKGRILNEKQAIINBIYSKSKSSNFSRTPPE
AKTRFEALIMKMHAEQGGILLIERRQQLKQV"
complement(8316. .10689)
/gene="T20F5.4"
/locus_tag="T20F5.4"
/locus_tag="T20F5.4"
/complement(join(8316. .8376,9000. .9220,9537. .9683,
9731. .9773,9826. .10047,10261. .10463,10585. .10689))
/gene="T20F5.4"
/locus_tag="T20F5.4"
/locus_tag="T20F5.4"
/standard_name="T20F5.4"
/notes="contains similarity to Candida glabrata Epa5p.;
TR:Q6B54; coded for by the following C. elegans cDNAs:
CK578651, CK580427, CK581704, OSTF079F11.1, OST079F11.1,
Yk109h3.3, Yk315c2.5, Yk394e10.3, Yk394e10.5, Yk477g6.3,
Yk477g6.5, Yk579h10.3, Yk579h10.5, Yk585B9.3, Yk585D9.5,
Yk600d6.3, Yk600d6.5, Yk605f7.3, Yk605f7.5, Yk681g5.3,
Yk681g5.5"
/codon_start=1
/product="Hypothetical protein T20F5.4"
/protein_id="AAB37665.1"
/db_xref="GI:1703617"
/db_xref=" WormBase: T20F5.4"
/translation="MTNMHDLRLNRRYACRSGMOELDMRLAKTLLIQLNLSRLPLI
FGSLDVOVREKTRDVLDFLRSLSSVSGSSQNTISTSTSEDEKTSNRHSNI
ENSSITMNVSTFRKKSLESTHSSVSHSNQNDYKTSFAPVYTSRSGRGMKTS
IRKPMQDSTTSNTVYIRSKPRPRHIDLTKERRSAFYLNQSSKHSIDELISSTLS
QSLNDEAKTPQKAIVIRKMKOKAMDAMPNAVYVYSSTEQIPKRAPLFSRARSPPV
PALPILKPTQHFAPCPKIRSPKNTKTVPAKVIAPVPSISPKVAKSTYLQSOPL
11555. .11882
/product="probable non-coding RNA"
/standard_name="standard name=T20F5.1"
/notes="coded for by the following cDNAs: OSTF069G4_1
OSTR069G4_1"
/db_xref=" WormBase: T20F5.1"
/complement(13927. .14664)
/gene="T20F5.5"
/locus_tag="T20F5.5"

```

RESULT 4				
AC161379				
LOCUS				
DEFINITION	AC161379	189293 bp	DNA	linear ROD 21-JUN-2005
ACCESSION	Mus musculus BAC clone RP23-207P12	from chromosome 17, complete sequence.		
VERSION	AC161379 AC120344			
KEYWORDS	AC161379.2 GI:66793614			
SOURCE	HTS.			
ORGANISM	Mus musculus (house mouse)			
REFERENCE	Mus musculus			
AUTHORS	Eukaryota; Metazoa; Chordata; Cranialia; Vertebrate; Euteleostomi;			
TITLE	Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;			
JOURNAL	Sciurognathi; Muridae; Murinae; Mus.			
REFERENCE	1 (bases 1 to 189293)			
AUTHORS	Lek, S., Kozlowski, A. and Haglund, K.			
TITLE	The sequence of Mus musculus BAC clone RP23-207P12			
JOURNAL	Unpublished (2001)			
REFERENCE	2 (bases 1 to 189293)			
AUTHORS	Wilson, R.K.			
TITLE	Direct Submission			
JOURNAL	Submitted (12-MAY-2005) Genome Sequencing Center, 4444 Forest Park			
REFERENCE	Parkway, St. Louis, MO 63108, USA			
AUTHORS	3 (bases 1 to 189293)			
TITLE	Wilson, R.K.			
JOURNAL	Direct Submission			
REFERENCE	Submitted (29-MAY-2005) Genome Sequencing Center, 4444 Forest Park			
AUTHORS	Parkway, St. Louis, MO 63108, USA			
TITLE	4 (bases 1 to 189293)			
JOURNAL	Wilson, R.K.			
REFERENCE	Direct Submission			
AUTHORS	Submitted (21-JUN-2005) Genome Sequencing Center, Washington			
TITLE	University School of Medicine, 4444 Forest Park Parkway, St. Louis,			
JOURNAL	MO 63108, USA			
COMMENT	On May 29, 2005 this sequence version replaced gi:63253526.			

----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu>
 Contact: submissions@watson.wustl.edu
 ----- Summary Statistics
 Center project name: M_BA0207P12
 Drafting center: WIBR

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e. phred quality >=30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone, fosmid clone or direct clone walk sequence. Sequence from the Mouse Genome Sequencing Consortium whole genome shotgun may have been used to obtain the consensus sequence. The assembly was confirmed by restriction digest. This finishing standard has slightly changed from the previous Human standard. Specifically, standards for regions of low sequence complexity (such as dinucleotide repeats and small unit tandem repeats) have been relaxed. These regions are very prevalent in the mouse genome, and the return on extended finishing efforts is minimal. If a sequence meets the criteria of the above statement, it needs no comments or tags. If the criteria are not met, such as ambiguous bases, then the region is duly annotated.

MAPPING INFORMATION: Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

SOURCE INFORMATION:

The BAC library has been constructed by Kazuhiro Osegawa and Minako Tateo in the laboratory of Pieter de Jong (<http://www.chori.org>) from female C57BL/6J mouse kidney and/or brain genomic DNA. The clone and detailed information can be obtained from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

This sequence is the entire insert of the clone.

FEATURES

source

Location/Qualifiers
1.189293
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="17"
/clone="RP23-207P12"
/clone_1ib="RP23-23"

ORIGIN

Query Match 44.7%; Score 28.6; DB 9; Length 189293;
Best Local Similarity 67.8%; Pred. No. 69;
Matches 40; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

3 TCCGGAAGAGGTCTCAAGCGCTGAGAGCTGTTAGCAAAATCTGAAGTGA 61
DB 185149 TCCAGAGATGCTGATGACCCGACAGAACTTGAGGAGAAATCAAGACAGCA 185207

RESULT 5
AC154577/ 201463 bp DNA linear HTG 30-DEC-2004
LOCUS Mus musculus chromosome 17 clone RP23-375G9, WORKING DRAFT
DEFINITION
SEQUENCE, 10 unordered pieces.
AC154577.1 GI:56900203

AC154577.1 GI:56900203
HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
Mus musculus (house mouse)

SOURCE

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 201463)

REFERENCE

Wilson, R.K.
The sequence of Mus musculus clone

AUTHORS

Unpublished

JOURNAL

2 (bases 1 to 201463)

REFERENCE

Wilson, R.K.

AUTHORS

Unpublished

JOURNAL

2 (bases 1 to 201463)

COMMENT

Submitted (30-DEC-2004) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
On Dec 30, 2004 this sequence version replaced gi:45356321.

Genome Center

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu>
Contact: submissions@wustl.edu
Project Information
Center project name: M_BA0375G09
Drafting center: WUGSC

Summary Statistics

Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-terminator Big Dye; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990119
Consensus quality: 196204 bases at least Q40
Consensus quality: 197587 bases at least Q30
Consensus quality: 198347 bases at least Q20

* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1271: contig of 1271 bp in length
* 1272 1371: gap of unknown length
* 1372 2636: contig of 1265 bp in length
* 2637 2737: gap of unknown length
* 2737 4059: contig of 1323 bp in length
* 4060 4159: gap of unknown length
* 4160 6162: contig of 2003 bp in length
* 6163 6263: gap of unknown length
* 6263 8067: contig of 1805 bp in length
* 8068 8168: gap of unknown length
* 8168 9486: contig of 1319 bp in length
* 9487 9586: gap of unknown length
* 9587 43296: contig of 33610 bp in length
* 43297 78014: gap of unknown length
* 78015 78115: gap of unknown length
* 78115 134599: contig of 56485 bp in length
* 134600 134699: gap of unknown length
* 134700 201463: contig of 66764 bp in length.

FEATURES

source

Location/Qualifiers
1.201463
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="17"
/clone="RP23-375G9"
1.1271

misc_feature

/note="assembly_name:Contig10"

gap

/estimated_length=unknown

misc_feature

/note="assembly_name:Contig11"

gap

/estimated_length=unknown

misc_feature

/note="assembly_name:Contig33"

gap

/estimated_length=unknown

misc_feature

/note="assembly_name:Contig34"

gap

/estimated_length=unknown

misc_feature

/note="assembly_name:Contig35"

gap

/estimated_length=unknown

misc_feature

/note="assembly_name:Contig37"

gap

/estimated_length=unknown

misc_feature

/note="assembly_name:Contig39"

gap

/estimated_length=unknown

misc_feature

/note="assembly_name:Contig40"

gap

/estimated_length=unknown

misc_feature

/note="assembly_name:Contig41"

gap

/estimated_length=unknown

misc_feature

/note="assembly_name:Contig42"

ORIGIN

Query Match 44.7%; Score 28.6; DB 14; Length 201463;
Best Local Similarity 67.8%; Pred. No. 70;

169293 169342: gap of 50 bp
169343 176726: contig of 784 bp in length
176727 176776: gap of 50 bp
176777 186821: contig of 1045 bp in length
186822 186871: gap of 50 bp
186872 197336: contig of 10465 bp in length
197337 197386: gap of 50 bp
197387 200928: contig of 3542 bp in length
200929 200978: gap of 50 bp
200979 213723: contig of 12745 bp in length
213724 213773: gap of 50 bp
213774 220586: contig of 6913 bp in length
220587 220736: gap of 50 bp
220737 241587: contig of 20851 bp in length
241588 241637: gap of 50 bp
241638 254209: contig of 12572 bp in length
254210 254259: gap of 50 bp
254260 259183: contig of 4924 bp in length
259184 259233: gap of 50 bp
259234 264358: contig of 5125 bp in length
264359 264458: gap of unknown length
264459 265533: contig of 1075 bp in length
265534 265633: gap of unknown length
265634 266654: contig of 1021 bp in length
266655 266754: gap of unknown length
266755 267755: contig of 1001 bp in length
267756 267855: gap of unknown length
267856 269273: contig of 1418 bp in length
269274 269373: gap of unknown length
269374 270505: contig of 1133 bp in length
270506 270605: gap of unknown length
270606 280474: contig of 9869 bp in length.

FEATURES

source 1. 280474
/organism="Bos taurus"
/mol_type="genomic DNA"
/db_xref="taxon:9913"
/clone="CH240-80023"
28680. 28729
/estimated_length=50
gap 35702. 35751
/estimated_length=50
gap 43865. 43914
/estimated_length=50
gap 56381. 56430
/estimated_length=50
gap 57988. 58037
/estimated_length=50
gap 76239. 76288
/estimated_length=50
gap 82248. 82297
/estimated_length=50
gap 112456. 112505
/estimated_length=50
gap 113990. 141935
/estimated_length=27946
gap 146587. 146636
/estimated_length=50
gap 159586. 159635
/estimated_length=50
gap 169293. 169342
/estimated_length=50
gap 176727. 176776

Query Match 44.7%; Score 28.6; DB 14; Length 280474;
Best Local Similarity 67.8%; Pred. No. 70;
Matches 40; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 2 ATCCGCAAGAGGTTCACAAAGCGCTGAGAGAGCTGTTAGCAAAATCTGGAAGTCTGA 60
DB 94280 ATCCATTGAGAGCTGCTCAAGAGCCTGGTGATGCTGTTTATCCAGTGTTCACCTGGA 94338

RESULT 7
AC006901
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL

AC006901 294136 bp DNA linear HTG 26-FEB-1999
Caenorhabditis elegans clone Y74A11X. *** SEQUENCING IN PROGRESS
***, 81 unordered pieces.
AC006901
AC006901.2 GI:4309901
HTG, HTGS, PHASE1.
Caenorhabditis elegans
Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditioidea; Rhabditidae; Peloderiinae; Caenorhabditis.
1 (bases 1 to 294136)
Waterston, R.H.
The sequence of Caenorhabditis elegans clone
Unpublished
2 (bases 1 to 294136)
Waterston, R.H.
Direct Submission
Submitted (24-FEB-1999) Genome Sequencing Center, Washington
University School of Medicine, 444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Mar 1, 1999 this sequence version replaced gi:4263447.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 81 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 2237 2236: contig of 2236 bp in length
2237 2251: gap of unknown length
2251 2252 4429: contig of 2178 bp in length
2252 4430 4444: gap of unknown length
4430 4445 6563: contig of 2119 bp in length
4445 6564 6578: gap of unknown length
6564 6579 8594: contig of 2016 bp in length
6579 8595 8609: gap of unknown length
8595 8610 10706: contig of 2097 bp in length
8610 10707 10721: gap of unknown length
10707 10722 12666: contig of 2145 bp in length
10722 12667 12881: gap of unknown length
12667 12882 15897: contig of 3016 bp in length
12882 15898 15912: gap of unknown length
15898 15913 18447: contig of 2535 bp in length
15913 18448 18462: gap of unknown length
18448 18463 21337: contig of 2875 bp in length
18463 21338 21352: gap of unknown length
21338 21353 24542: contig of 3190 bp in length
21353 24543 24557: gap of unknown length
24543 24558 26812: contig of 2255 bp in length
24558 26813 29303: gap of unknown length
26813 29304 29318: gap of unknown length
29304 29319 31459: gap of 2141 bp in length
29319 31460 31474: gap of unknown length
31460 31475 33742: contig of 2268 bp in length
31475 33743 33757: gap of unknown length
33743 33758 35945: contig of 2188 bp in length
33758 35946 35960: gap of unknown length
35946 35961 39232: contig of 3272 bp in length
35961 39233 39247: gap of unknown length
39233 39248 42387: contig of 3140 bp in length
39248 42388 42402: gap of unknown length
42388 42403 42419: contig of 2797 bp in length
42403 42404 45200: gap of unknown length
42404 45201 47602: contig of 2388 bp in length
45201 47603 47617: gap of unknown length
47603 47618 50110: contig of 2493 bp in length
47618 50111 52939: gap of unknown length
50111 52940 52954: contig of 2814 bp in length
52939 52940 52954: gap of unknown length

*	52965	55025:	contig of 2071	bp in length
*	55026	55040:	gap of unknown	length
*	55041	57657:	contig of 2617	bp in length
*	57658	57672:	gap of unknown	length
*	57673	60530:	contig of 2858	bp in length
*	60531	60545:	gap of unknown	length
*	60546	64221:	contig of 3676	bp in length
*	64222	64236:	gap of unknown	length
*	64237	66744:	contig of 2508	bp in length
*	66745	66759:	gap of unknown	length
*	66760	69537:	contig of 2178	bp in length
*	69538	69552:	gap of unknown	length
*	69553	71778:	contig of 2226	bp in length
*	71779	71793:	gap of unknown	length
*	71794	73966:	contig of 2173	bp in length
*	73967	73981:	gap of unknown	length
*	73982	75966:	contig of 1985	bp in length
*	75967	75981:	gap of unknown	length
*	75982	78741:	contig of 2760	bp in length
*	78742	78756:	gap of unknown	length
*	78757	81353:	contig of 2602	bp in length
*	81359	81373:	gap of unknown	length
*	81374	84070:	contig of 2697	bp in length
*	84071	84085:	gap of unknown	length
*	84086	86914:	contig of 2829	bp in length
*	86915	86929:	gap of unknown	length
*	86930	88248:	contig of 2319	bp in length
*	89249	89263:	gap of unknown	length
*	89249	91256:	contig of 1993	bp in length
*	91257	91271:	gap of unknown	length
*	91272	95588:	contig of 4317	bp in length
*	95589	96003:	gap of unknown	length
*	95604	97920:	contig of 2317	bp in length
*	97921	97935:	gap of unknown	length
*	97936	100805:	contig of 2870	bp in length
*	100806	100820:	gap of unknown	length
*	100821	104770:	contig of 3950	bp in length
*	104771	104785:	gap of unknown	length
*	104786	107676:	contig of 2891	bp in length
*	107677	107691:	gap of unknown	length
*	107692	110889:	contig of 3198	bp in length
*	110890	110904:	gap of unknown	length
*	110905	111390:	contig of 2486	bp in length
*	111391	111405:	gap of unknown	length
*	111406	116553:	contig of 3148	bp in length
*	116554	116568:	gap of unknown	length
*	116569	119084:	contig of 2516	bp in length
*	119085	119099:	gap of unknown	length
*	119100	122456:	contig of 3357	bp in length
*	122457	122471:	gap of unknown	length
*	122472	125606:	contig of 3135	bp in length
*	125607	125621:	gap of unknown	length
*	125622	128558:	contig of 2937	bp in length
*	128559	128673:	gap of unknown	length
*	128574	132627:	contig of 4054	bp in length
*	132628	132842:	gap of unknown	length
*	132643	135451:	contig of 2809	bp in length
*	135452	135466:	gap of unknown	length
*	135467	138584:	contig of 3118	bp in length
*	138585	138599:	gap of unknown	length
*	138600	141494:	contig of 2895	bp in length
*	141495	141509:	gap of unknown	length
*	141510	145190:	contig of 3681	bp in length
*	145191	145205:	gap of unknown	length
*	145206	148105:	contig of 2900	bp in length
*	148106	148120:	gap of unknown	length
*	148121	152177:	contig of 4057	bp in length
*	152178	152192:	gap of unknown	length
*	152193	155494:	contig of 3057	bp in length
*	155230	155264:	gap of unknown	length
*	155265	159702:	contig of 4438	bp in length
*	159703	159717:	gap of unknown	length
*	159718	163556:	contig of 4139	bp in length

	Query Match	Best Local Similarity	Mismatches	Gaps
OY	14 GGTTGCAACGCCCTGGAGAGCTGTATTACAAATCTGGAAGTGAAGA 64	44.7% Score 28.6; DB 14; Length 294136;	0;	0;
Dn	148891 GGTCTCACAACACTGTAGAAAGTGTCAGACAGAAATCTGTGTGAAGA 148941	72.5% Pred. No. 71; Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;		
RESULT 8				
ACI64240/c				
LOCUS	ACI64240	172866 bp	DNA	linear HTG 01-JUN-2005
DEFINITION	Bos taurus clone CH240-138D7, *** SEQUENCING IN PROGRESS ***, 13			
ACCSSION	ACI64240			
VERSION	ACI64240.2 GI:68301068			

KEYWORDS
SOURCE
ORGANISM
AUTHORS

REFERENCE

HTG, HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
Bos taurus (cow)
Bos taurus
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
Peccora; Bovidae; Bovinae; Bos.
1 (bases 1 to 172866)

Muzny, D., Marie, M., Metzger, M., Lee, A., Abramo, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benhammed, F., Biewald, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinu, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Frazer, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Geisberg, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Haves, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogue, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowals, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensueta, L., Louleseg, H., Lozano, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Ming, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Muniasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwankwelu, O., Okunolu, G., Olarunpusagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankoch, C., Plopper, F., Polindexter, A., Popovic, D., Primus, E., Pu, L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, M., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartbeyn, A., Sisson, I., Sitter, C. D., Sma's, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soza, J., Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Uman, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willison, R., Wlezyk, R., Woodson, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G., and Gibbs, R. A.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Submitted (18-JUN-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 172866)
Direct Submission
Cow Genome Sequencing Consortium.
Submitted (01-JUL-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jun 29, 2005 this sequence version replaced gi:67972549.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas

(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

Project Information
Center project name: FTMG
Center clone name: CH240-13807

Assembly program: Atlas 3.0;
Consensus quality: 16797 bases at least Q40
Consensus quality: 169570 bases at least Q20
Consensus quality: 170971 bases at least Q20
Estimated insert size: 169477; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
NOTE: This is a 'working draft' sequence. It currently consists of 13 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1	9129: contig of 9129 bp in length
9130	9179: gap of 50 bp
9180	55666: contig of 46467 bp in length
55647	55696: gap of 50 bp
55697	64790: contig of 9094 bp in length
64791	64840: gap of 50 bp
64841	96307: contig of 31467 bp in length
96308	96357: gap of 50 bp
96358	117107: contig of 20750 bp in length
117108	117324: gap of 217 bp
117325	118341: contig of 1017 bp in length
118342	118391: gap of 50 bp
118392	151210: contig of 32819 bp in length
151211	151260: gap of 50 bp
151261	163474: contig of 12214 bp in length
163475	163524: gap of 50 bp
163525	166739: contig of 3215 bp in length
166740	166839: gap of unknown length
166840	168328: contig of 1489 bp in length
168329	168428: gap of unknown length
168429	169601: contig of 1173 bp in length
169602	169701: gap of unknown length
169702	171293: contig of 1592 bp in length
171294	171393: gap of unknown length
171394	172866: contig of 1473 bp in length.

FEATURES
source

1. 172866
/organism="Bos taurus"
/mol_type="genomic DNA"
/db_xref="taxon:9913"
/clone="CH240-13807"
9130..9179
/estimated_length=50
55647..55696
/estimated_length=50
64791..64840
/estimated_length=50

gap 96308..96357
 /estimated_length=50
 gap 117108..117324
 /estimated_length=217
 gap 118342..118391
 /estimated_length=50
 gap 151211..151260
 /estimated_length=50
 gap 163475..163524
 /estimated_length=50
 gap 166740..166839
 /estimated_length=unknown
 gap 168329..168428
 /estimated_length=unknown
 gap 169602..169701
 /estimated_length=unknown
 gap 171294..171393
 /estimated_length=unknown

ORIGIN

Query Match 44.4%; Score 28.4; DB 14; length 172866;
 Best Local Similarity 70.4%; Pred. No. 81;
 Matches 38; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Oy 10 AGAAGGTCTCAACGCTGAGAGAGCTGTTAGCAAAATCGAAGCTGGAAGT 63
 Db 146110 ACAAGGTCTCCAGGCTGTAGAGCCATTCATCAGCTCGCAGCTGAGGT 146057

RESULT 9

AC149957

LOCUS

DEFINITION

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

AC149957

COMMENT

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

FEATURES

Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G.,
 Scherer, S., Scott, G., Shen, H., Shoshitani, N., Sisson, I.,
 Sodergren, E., Sonaike, A., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Swalek, A., Tabor, P., Tamersia, A., Tamersia, K., Tang, H.,
 Tansley, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Uemami, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wellington, S.,
 Williams, G., Williamson, A., Wleczyk, R., Woodson, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G. and Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 145870)
 Worley, K. C.
 Direct Submission
 Submitted (25-JUN-2004) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 ----- Genome Center -----
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information -----
 Center project name: SPCO
 Center clone name: R3-32B13
 ----- Summary Statistics -----
 Sequencing vector: plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 134633 bases at least Q40
 Consensus quality: 136701 bases at least Q30
 Consensus quality: 138522 bases at least Q20
 Estimated insert size: 140142; sum-of-coverage estimation
 Quality coverage: 4x in Q20 bases; sum-of-coverage estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 13 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 2075: contig of 2075 bp in length
 * 2076 2173: gap of unknown length
 * 2176 3639: contig of 1464 bp in length
 * 3640 3739: gap of unknown length
 * 3740 4912: contig of 1173 bp in length
 * 4913 5012: gap of unknown length
 * 5013 6092: contig of 1080 bp in length
 * 6093 6192: gap of unknown length
 * 6193 9788: contig of 3596 bp in length
 * 9789 9888: gap of unknown length
 * 9889 11691: contig of 1803 bp in length
 * 11692 11791: gap of unknown length
 * 11792 21915: contig of 10124 bp in length
 * 21916 22015: gap of unknown length
 * 22016 31447: contig of 9432 bp in length
 * 31448 31547: gap of unknown length
 * 31548 46708: contig of 15161 bp in length
 * 46709 46808: gap of unknown length
 * 46809 59760: gap of 12952 bp in length
 * 59761 59860: gap of unknown length
 * 59861 84333: contig of 24473 bp in length
 * 84334 84433: gap of unknown length
 * 84434 112037: contig of 27604 bp in length
 * 112038 112138: gap of unknown length
 * 112139 145870: contig of 33733 bp in length.
 Location/Qualifiers
 1. 145870

```

/organism="Strongylocentrotus purpuratus"
/mol_type="genomic DNA"
/db_xref="taxon:7668"
/clone="R3-32B13"
2076.2175
/estimated_length=unknown
3640.3739
/estimated_length=unknown
4913.5012
/estimated_length=unknown
6093.6192
/estimated_length=unknown
9789.9988
/estimated_length=unknown
11692.11791
/estimated_length=unknown
21916.22015
/estimated_length=unknown
31448.31547
/estimated_length=unknown
46709.46808
/estimated_length=unknown
59761.59860
/estimated_length=unknown
84334.84433
/estimated_length=unknown
112038.112137
/estimated_length=unknown

```

ORIGIN

```

Query Match      44.1%; Score 28.2; DB 14; Length 145870;
Best Local Similarity 68.4%; Pred. No. 95;
Matches 39; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

```

```

QY      6 GCGAAGAGGTGTTCAAGCGCTGAGAGCTTTAGCAAAATCTGGAAGTGAAG 62
Db      128556 GCCAAGACGAGTTGACGCAAGAAAGTGAAGCTATAGATTAGTGAAGTGAAG 128612

```

```

RESULT 10      229461 bp      DNA      linear      HTG 01-JUL-2005
AC155077/c     Bos taurus clone CH240-35K10.*** SEQUENCING IN PROGRESS ***. 22
LOCUS
DEFINITION
unordered pieces.

```

```

AC155077
AC155077.2 GI:68265207
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE
Bos taurus (cow)

```

```

ORGANISM
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
Pecora; Bovidae; Bovinae; Bos.
1 (bases 1 to 229461)

```

REFERENCE

AUTHORS

Wozny, D., Martie, Metzger, M., Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biewald, K., Blair, J., Blankenburg, K., Blych, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Butrell, K., Calderon, E., Cardenas, V., Carter, K., Cavaros, I., Cesar, H., Center, A., Chacho, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falle, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Gante, R., Garcia, A., Garner, T., Garza, M., Gebreyes, E., Geer, A., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamli, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Huily, S., Hume, J., Idlebird, D., Jackson, A.,

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowik, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshaw, L., Louiseged, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindratne, M., Mahmoud, M., Mallory, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mahoney, S., McLeod, M.P., McNeill, T.Z., Meenen, B., Milosavljevic, A., Miner, G., Minja, E., Montemayor, D., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K., Nwackelme, O., Okwum, G., Olariu, A., Olariu, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Prims, B., Pu, L., Puzo, M., Quirio, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, R., Rose, R., Ruiz, S., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tinney, A., Trejos, Z., Umami, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, P., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, S., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstein, G., and Gibbs, R.A.

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

2 (bases 1 to 229461)
 Worley, K.C. 2005 Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 229461)
 Cow Genome Sequencing Consortium.
 Direct Submission
 Submitted (01-JUL-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On Jun 28, 2005 this sequence version replaced gi:57334865.
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu

----- Project Information
 Center project name: FC12
 Center clone name: CH240-35K10

----- Summary Statistics
 Assembly program: Atlas 3.0
 Consensus quality: 218943 bases at least Q40
 Consensus quality: 221244 bases at least Q30
 Consensus quality: 223215 bases at least Q20
 Estimated insert size: 222266; sum-of-contigs estimation
 Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length

```
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html)
* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
```

*	1	42348:	contig of 42348 bp in length
*		42349	gap of 50 bp
*		42399	50935: contig of 8537 bp in length
*		50936	50985: gap of 50 bp
*		50986	63009: contig of 12024 bp in length
*		63010	63644: gap of 635 bp
*		63645	69641: contig of 5997 bp in length
*		69642	69849: gap of 208 bp
*		69850	79590: contig of 9741 bp in length
*		79591	80666: gap of 1076 bp
*		80667	83794: contig of 3128 bp in length
*		83795	83894: gap of unknown length
*		83895	86322: contig of 2428 bp in length
*		86323	86601: gap of 279 bp
*		86602	88815: contig of 2214 bp in length
*		88816	88915: gap of unknown length
*		88916	97355: contig of 8440 bp in length
*		97356	97405: gap of 50 bp
*		97406	137450: contig of 40045 bp in length
*		137451	137550: gap of unknown length
*		137551	144447: contig of 6897 bp in length
*		144448	144497: gap of 50 bp
*		14498	200363: contig of 53666 bp in length
*		200364	200413: gap of 50 bp
*		200414	202340: contig of 1927 bp in length
*		202341	202390: gap of 50 bp
*		202391	204245: contig of 1855 bp in length
*		204246	204345: gap of unknown length
*		204346	205570: contig of 1225 bp in length
*		205571	205670: gap of unknown length
*		205671	206913: contig of 1243 bp in length
*		206914	207013: gap of unknown length
*		207014	208085: contig of 1072 bp in length
*		208086	208185: gap of unknown length
*		208186	209217: contig of 1032 bp in length
*		209218	209317: gap of unknown length
*		209318	211040: contig of 1723 bp in length
*		211041	211140: gap of unknown length
*		211141	212172: contig of 1032 bp in length
*		212173	212272: gap of unknown length
*		212273	213321: contig of 1049 bp in length
*		213322	213421: gap of unknown length
*		213422	229461: contig of 16040 bp in length.

```
Location/Qualifiers
1..229461
/organism="Bos taurus"
/mol_type="genomic DNA"
/db_xref="taxon:9913"
/clone="CH240-35K10"
42349..42398
/estimated_length=50
50936..50985
/estimated_length=50
63010..63644
/estimated_length=635
69642..69849
/estimated_length=208
79591..80666
/estimated_length=1076
83795..83894
/estimated_length=unknown
86323..86601
/estimated_length=279
88816..88915
```

gap	/estimated_length=unknown	97356. .97405	/estimated_length=50	137451. .137550	/estimated_length=unknown	144448. .144497	/estimated_length=50	200364. .200413	/estimated_length=50	202341. .202390	/estimated_length=50	204246. .204345	/estimated_length=unknown	205571. .205670	/estimated_length=unknown	206914. .207013	/estimated_length=unknown	208086. .208185	/estimated_length=unknown	209218. .209317	/estimated_length=unknown	211041. .211140	gap
Query Match	44.1%;	Score 28.2;	DB 14;	Length 229461;																			
Best Local Similarity	73.5%;	Pred. No. 96;																					
Matches	36;	Conservative	0;	Mismatches	13;	Indels	0;	Gaps	0;														
Qy	16	TGTTCAACGCGCTGGAGACGCTGTTTACCAAAATCTGGAACTGGAAGTA	64																				
Db	191647	TGTTCAATCAGCTGGAGAAATTTGATGACGACGATCAGGGCCTGGAAAGTA	191599																				
RESULT 11	AC125238	196622 bp	DNA	linear	PRI 30-APR-2005																		
LOCUS	Homo sapiens BAC clone Rpl1-711C24	from 2, complete sequence.																					
DEFINITION	AC125238																						
ACCESSION	AC125238.5	GI:22267884																					
VERSION	HTG.																						
KEYWORDS	Homo sapiens (human)																						
SOURCE	Homo sapiens																						
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;																						
	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;																						
	Homnidae; Homo.																						
REFERENCE	1 (bases 1 to 196622)																						
AUTHORS	Nguyen,C., Bielicki,L. and Schatzkamer,K.																						
TITLE	The sequence of Homo sapiens BAC clone Rpl1-711C24																						
JOURNAL	Unpublished (2001)																						
REFERENCE	2 (bases 1 to 196622)																						
AUTHORS	Waterston,R.H.																						
TITLE	Direct Submission																						
JOURNAL	Submitted (20-JUN-2002) Genome Sequencing Center, Washington																						
	University School of Medicine, 4444 Forest Park Parkway, St. Louis,																						
	MO 63108, USA																						
REFERENCE	3 (bases 1 to 196622)																						
AUTHORS	Waterston,R.H.																						
TITLE	Direct Submission																						
JOURNAL	Submitted (08-AUG-2002) Genome Sequencing Center, Washington																						
	University School of Medicine, 4444 Forest Park Parkway, St. Louis,																						
	MO 63108, USA																						
REFERENCE	4 (bases 1 to 196622)																						
AUTHORS	Waterston,R.																						
TITLE	Direct Submission																						

JOURNAL

COMMENT

Submitted (30-APR-2005) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
 On Aug 16, 2002 this sequence version replaced gi:22338709.

----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu>
 Contact: submissions@wustl.wustl.edu
 Summary Statistics
 Center project name: H_NH0711C24

NOTICE:

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

SOURCE INFORMATION:

The RPCT-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Moon, P.Y., Zhao, B., Frengen, E., Tateo, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.edu>
 VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone. This clone is overlapped by AC069314 and AC010138.

Location/Qualifiers

1. 196622

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="2"

/clone_1b="RPCT-11"

/clone_1lb="RPCT-11"

/note="CpG island (%GC=60.7, o/e=0.67, #CpGs=18)"

/note="CpG island (%GC=57.4, o/e=1.17, #CpGs=21)"

/note="CpG island (%GC=73.5, o/e=0.70, #CpGs=31)"

/note="CD28"

/note="CD28"

/note="CD28"

/note="CD28"

/note="CD28"

/note="CD28"

/note="CD28"

/note="CD28"

/note="CD28"

/note="CD28"

/note="CD28"

/note="CD28"

/note="CD28"

/note="CD28"

/note="CD28"

/note="CD28"

/note="CD28"

/note="CD28"

/note="CD28"

/note="CD28"

/note="CD28"

/note="CD28"

FEATURES

source

1. 196622

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="2"

/clone_1b="RPCT-11"

/clone_1lb="RPCT-11"

/note="CpG island (%GC=60.7, o/e=0.67, #CpGs=18)"

/note="CpG island (%GC=57.4, o/e=1.17, #CpGs=21)"

/note="CpG island (%GC=73.5, o/e=0.70, #CpGs=31)"

/note="CD28"

/note="CD28"

/note="CD28"

/note="CD28"

/note="CD28"

/note="CD28"

/note="CD28"

/note="CD28"

/note="CD28"

/note="CD28"

/note="CD28"

/note="CD28"

/note="CD28"

/note="CD28"

/note="CD28"

/note="CD28"

/note="CD28"

/note="CD28"

/note="CD28"

/note="CD28"

/note="CD28"

ORIGIN

Query Match 43.4%; Score 27.8; DB 8; Length 196622;
 Best Local Similarity 69.1%; Pred. No. 1.3e+02;
 Matches 38; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

1 GATCCGCGAGAGGCTGTTCAACGCCCTGAGAGCTTTAGCAAAATCTGAA 55
 Db 157692 GAACTGAGAGCTGAGAGCAAAATCTGAGAGAGCAAAATCTGGA 157638

RESULT 12

HSU78045

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

YVNGTDIYFKIEVWPPYLLDNEKSGNTIIHVKGKILCBSPLEFGSPKFWLVVVG
 GVLAQYSLVTVARIIFWVRSKRSRLHSDYMMNTPRRPGTRKHQYAPYAPRPAAY
 RS"

171163. 171427
 /note="CpG island (%GC=62.3, o/e=0.80, #CpGs=20)"
 177943. 178329
 /note="CpG island (%GC=73.1, o/e=0.60, #CpGs=34)"

43.4%; Score 27.8; DB 8; Length 196622;
 Best Local Similarity 69.1%; Pred. No. 1.3e+02;
 Matches 38; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

1 GATCCGCGAGAGGCTGTTCAACGCCCTGAGAGCTTTAGCAAAATCTGAA 55
 Db 157692 GAACTGAGAGCTGAGAGCAAAATCTGAGAGAGCAAAATCTGGA 157638

HSU78045 81826 bp DNA linear PRI 21-SEP-2001
 Human collagenase and stromelysin genes, complete cds, and
 metalloelastase gene, partial cds.
 U78045
 U78045.1 GI:1688257
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homidae; Homo.
 1 (bases 4225 to 11906)
 Goldberg, G.I., Wilhelm, S.M., Kronberger, A., Bauer, E.A., Grant, G.A.
 and Eisen, A.Z.
 Human fibroblast collagenase. Complete primary structure and
 homology to an oncogene transformation-induced rat protein
 J. Biol. Chem. 261 (14), 6600-6605 (1986)
 3009463
 2 (bases 49921 to 57338)
 Saus, J., Quinones, S., Orlan, Y., Nagase, H., Harris, E.D. Jr. and
 Kurkinen, M.
 The complete primary structure of human matrix metalloproteinase-3.
 Identity with stromelysin
 J. Biol. Chem. 263 (14), 6742-6745 (1988)
 3360803
 3 (bases 1 to 81826)
 Shapiro, S.D., Kobayashi, D.K. and Ley, T.J.
 Cloning and characterization of a unique elastolytic
 metalloproteinase produced by human alveolar macrophages
 J. Biol. Chem. 268 (32), 23824-23829 (1993)
 8286919
 4 (bases 1 to 81826)
 Borden, P. and Heller, R.A.
 Transcriptional control of matrix metalloproteinases and the tissue
 inhibitors of matrix metalloproteinases
 Crit. Rev. Eukaryot. Gene Expr. 7 (1-2), 159-178 (1997)
 9034720
 5 (bases 1 to 81826)
 Lin, D., Duncan, M., Allen, E., Araujo, R., Aparicio, A., Chai, A.,
 Chung, E., Davis, K., Federspiel, N., Hyman, R., Kaiman, S., Komp, C.,
 Kurd, O., Lashkari, D., Lew, H., Namath, A., Oefner, P., Roberts, D.,
 Heller, R. and Davis, R.W.
 Three Matrix Metalloproteinases on 81kb of P1 insert
 Unpublished
 6 (bases 1 to 81826)
 Lin, D.
 Direct Submission
 Submitted (12-NOV-1996) Biochemistry, Stanford University, 855
 California Ave, Palo Alto, CA 94304
 Location/Qualifiers
 1. 81826
 /organism="Homo sapiens"
 /mol_type="genomic DNA"

```

CDS
/db xref="taxon:9606"
complement(join(4225..4334,4516..4619,5146..5308,
6418..6551,8988..9105,9266..9421,10478..10603,
10828..10976,11070..11314,11801..11905))
/function="degrades fibrillar collagen"
/note="similar to human skin collagenase encoded by
Genbank Accession Number M1509"
/codon_start=1
/product="collagenase"
/protein_id="AAB36941.1"
/db_xref="GI:1688258"
/translation="MHSPPLLLLLFMGVSHSPATLETQEDVDLVQKYLEKYNL
KDGQVRRKNSGVEEKLKOMDEFGLKYGKDAETLKMQRGCVDAQVIL
TEGNPRMECHTLTYIENYTPDLPRADVDHAIKXAPOLMSNTPLTFKVSQGDIM
ISPVRGDRNSPDGPGCNLAHAPRGPGIGADHPDEDMNTNPREVNLHVAH
ELHSLGSHSTIDICALMPTSTSGDQQLADDIDGICQATYGRQNVQPIGQTPK
ACDSLTGTDATTTIRGEVWFEDRPRYMTNPPPEVLENFISVEPQIPLNGLEAVER
ADREVRFPKGNKYWAVQGVNLAHGYPDIVSSFGPPTVKHIDALAEENTGKTYFF
VANKYRYDEYKRSMDPGVPMKIAHDFPGIGHKVDAVFMKDGFEFFHGTQYKFDPK
TKRILTLQKANSWFCRKN"

misc_difference
9056
/note="when compared to Genbank Accession Number X54925"
/replace="t"
complement(join(11970..11973,12037..12045,12054..12063,
12208..12217))
/note="collagenase promoter"
/citation=(4)
complement(join(49921..50021,51093..51196,52346..52505,
52905..53038,53903..54047,54224..54388,55558..56083,
56222..56370,56463..56707,57234..57338))
/function="degrades matrix proteoglycans"
/note="similar to human matrix metalloproteinase-3 encoded
by Genbank Accession Number J03209"
/codon_start=1
/product="stromelysin"
/protein_id="AAB36942.1"
/db_xref="GI:1688259"
/translation="MKSLPILLILCVAVSAPYLDAARGEDTSNMLVOKLYENYDL
EKDVQAPRRKDSGVYKIKREMOKFLGIEVTKGKDSLTLEVMKPRGCVDVGHFRF
FPGIKRMRKHLTRYIVNTPDLPRDADSAVEKALKWEEETPLRFRLEGEADIM
ISFARHGDEYFPGPGNVLAHAAYAPGPGINGDAHPDDDEMTYDTGTGLFLVAH
EIGHSLGFHSANTALMYPLVHSLTDLRFSLSCDDINGISLGPSPDEPTELVV
TEPVPEPGTPANCDPALSPFADSLTDLRGELILFKRHFMRKSLRLBELHLISFVF
SLPSGVDAAYEVTSKDIIVPIKGNQPMALRGVAVAGYPRGHTIGPPTPKYDAAI
SDKEKNTYFVEDKYMRPDEKRNSEMGFPQIADPEPGIDSKIDAVFERGFFYFF
TGSQLEFDPNKKVTHLTKSNWJLNC"

misc_difference
56526
/note="when compared to Genbank Accession Number X05232"
/replace="a"
complement(join(76889..76989,77935..78041,79565..79724,
80104..80237,81059..81182,81697..81793))
/function="degrades elastin fibers"
/note="MTE: similar to human metalloelastase precursor
(matrix metalloproteinase-12), Swiss-Prot Accession Number
P39900"
/codon_start=1
/product="metalloelastase"
/protein_id="AAB36943.1"
/db_xref="GI:1688260"
/translation="DPKAVMPTKYKVDINTRLSADDIRGQISYGPKEKQRLPNP
DNSXPALCDPNLSFPAVTVGNKIFPFKDRFVWLVSERPKTSVNLISLWPTLPSGI
EAAVEIARNQVLEFKDKYWLISNLRPEPNPKSHSGFNPVKIKIDAVENRFY
RTYFVDQYMRIDRRQMDPGPKLITKNGQIGKPIDAVFYSKNKYVYFQGSNQ
FEYDILQRITKLSNWFQGC"

misc_difference
77945
/note="when compared to Genbank Accession Number L23808"
/replace="a"
misc_difference
81755
/note="when compared to Genbank Accession Number L23808"
/replace="g"

```

```

misc_difference 81782
/note="when compared to Genbank Accession Number L23808"
/replace="a"

ORIGIN
Query Match 42.8%; Score 27.4; DB 8; Length 81826;
Best Local Similarity 69.8%; Pred. No. 1.8e+02;
Matches 37; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Cy 3 TCCCGAAGAGGTCTTCAACCGCTGAGAGCTGTTTACCAAAATCTGAA 55
Db 6519 TCCGGATGAGAGGATTTGCGCATGTAGATGTATGAGAAATPAGCAA 6571

RESULT 13
AP000922/c
LOCUS
DEFINITION
Homo sapiens chromosome 11 clone RP11-75095 map 11q22, WORKING
DRAFT SEQUENCE, 35 unordered pieces.
ACCESSION
AP000922
VERSION
AP000922.2 GI:8119063
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
1 (bases 1 to 173359)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 173,359 genomic DNA of 11q22
Published Only in Database (1999)
2 (bases 1 to 173359)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (17-DEC-1999) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
Japan (E-mail:hattori@gsc.riken.go.jp,
URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
Fax:81-42-778-9924)
On May 31, 2000 this sequence version replaced gi:6997738.

----- Genome Center
Center: RIKEN Genomic Sciences Center (GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name: HumDrafc11
Center clone name: RP11-75095
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 149299 bases at least Q40
Consensus quality: 161098 bases at least Q30
Consensus quality: 167078 bases at least Q20
Insert size: 169959; sum-of-coverage
Quality coverage: 4.01x in Q20 bases; sum-of-coverage
-----
NOTE: This is a 'working draft' sequence. It currently consists of
35 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs 'N', but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved
1 15727 contig of 15727 bp in length
15828 27186 contig of 11359 bp in length
27287 37387 contig of 10101 bp in length
37488 47288 contig of 9801 bp in length
47389 55992 contig of 8604 bp in length

```

```

56093      63632 contig of 7540 bp in length
63733      73241 contig of 9509 bp in length
73342      80659 contig of 7318 bp in length
80760      89061 contig of 8302 bp in length
89162      94975 contig of 5814 bp in length
95076      100459 contig of 5384 bp in length
100560     106303 contig of 5744 bp in length
106404     111824 contig of 5421 bp in length
111825     111924 contig of 100 bp in length
111925     117598 contig of 5674 bp in length
117599     123774 contig of 6076 bp in length
123775     128021 contig of 4147 bp in length
128021     128122 contig of 4147 bp in length
128122     134542 contig of 4320 bp in length
134543     138972 contig of 4265 bp in length
138973     143337 contig of 4265 bp in length
143338     143438 contig of 3283 bp in length
143438     146720 contig of 2191 bp in length
146721     146820 contig of 2191 bp in length
146821     149011 contig of 2191 bp in length
149012     149111 contig of 2021 bp in length
149112     151332 contig of 2021 bp in length
151333     151233 contig of 2064 bp in length
151233     154504 contig of 2034 bp in length
154504     156567 contig of 1931 bp in length
156568     158701 contig of 1931 bp in length
158702     160732 contig of 1931 bp in length
160733     162522 contig of 1690 bp in length
162523     164124 contig of 1502 bp in length
164125     164225 contig of 1502 bp in length
164225     165574 contig of 1502 bp in length
165575     167245 contig of 1571 bp in length
167246     168587 contig of 1571 bp in length
168588     169577 contig of 1242 bp in length
169578     170997 contig of 1070 bp in length
170998     172148 contig of 1140 bp in length
172149     173359 contig of 1051 bp in length
173360     173359 contig of 1111 bp in length

Sequence updated (26-May-2000).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 35 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
15727: contig of 15727 bp in length
15728      15827: gap of 100 bp
15828      27186: contig of 11359 bp in length
27187      27286: gap of 100 bp
27287      37387: contig of 10101 bp in length
37388      37487: gap of 100 bp
37488      47288: contig of 9801 bp in length
47289      47388: gap of 100 bp
47389      55922: contig of 8604 bp in length
55930      56092: gap of 100 bp
56093      63632: contig of 7540 bp in length
63633      63732: gap of 100 bp
63733      73241: contig of 9509 bp in length
73242      73341: gap of 100 bp
73342      80659: contig of 7318 bp in length
80660      80759: gap of 100 bp
80760      89061: contig of 8302 bp in length
89062      89161: gap of 100 bp
89162      94975: contig of 5814 bp in length
94976      95076: gap of 100 bp
95077      100459: contig of 5384 bp in length
100460     100559: gap of 100 bp
100560     106303: contig of 5744 bp in length
106304     106403: gap of 100 bp
106404     111824: contig of 5421 bp in length
111825     111924: gap of 100 bp
111925     117598: contig of 5674 bp in length
117599     123774: contig of 6076 bp in length
123775     128021: contig of 4147 bp in length
128022     128122: gap of 100 bp
128123     134542: contig of 4320 bp in length
134543     138972: gap of 100 bp

```

```

* 134643      138972: contig of 4330 bp in length
* 138973      139072: gap of 100 bp
* 139073      143337: contig of 4265 bp in length
* 143338      143437: gap of 100 bp
* 143438      146720: contig of 3283 bp in length
* 146721      146820: gap of 100 bp
* 146821      149011: contig of 2191 bp in length
* 149012      149111: gap of 100 bp
* 149112      151332: contig of 2021 bp in length
* 151333      151233: gap of 100 bp
* 151233      154504: contig of 3171 bp in length
* 154504      156567: contig of 2064 bp in length
* 156568      158701: contig of 2034 bp in length
* 158702      158801: gap of 100 bp
* 158802      160732: contig of 1931 bp in length
* 160733      162522: contig of 1690 bp in length
* 162523      164124: contig of 1502 bp in length
* 164125      164225: gap of 100 bp
* 164225      165574: contig of 1350 bp in length
* 165575      167245: contig of 1571 bp in length
* 167246      168587: contig of 1571 bp in length
* 168588      169577: contig of 1242 bp in length
* 169578      170997: contig of 1070 bp in length
* 170998      172148: contig of 1140 bp in length
* 172149      173359: contig of 1051 bp in length
* 173360      173359: contig of 1111 bp in length

FEATURES
source
1. 173359
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q22"
/clone="RP11-750P5"
1. 15727
/note="assembly_fragment"
15828..27186
/note="assembly_fragment clone_end:SP6 vector_side:right"
27287..37387
/note="assembly_fragment"
37488..47288
/note="assembly_fragment"
47389..55922
/note="assembly_fragment"
56093..63632
/note="assembly_fragment"
63733..73241
/note="assembly_fragment"
73342..80659
/note="assembly_fragment"
80760..89061
/note="assembly_fragment clone_end:IT7 vector_side:right"
89162..94975
/note="assembly_fragment"
95076..100459
/note="assembly_fragment"
100560..106303
/note="assembly_fragment"
106404..111824
/note="assembly_fragment"
111925..117598
/note="assembly_fragment"

```

Query Match 42.8%; Score 27.4; DB 14; Length 173359;
Best Local Similarity 69.8%; Pred. No. 1.8e+02;

Matches	377	Conservative	0	Mismatches	16	Indels	0	Gaps	0
Qy	3	TCGCGGAAGAAGGTGTTCAAAAGCCTGAGAGAAGCTGTTTACGAAATTCGAA	55						
Db	91412	TCGCGGTGAAGAGGATTGTGCGCATGTAGAAATCTGAATTAGAAAATTAAGCAA	91360						
RESULT 14									
AP001144/c									
LOCUS									
DEFINITION	AP001144	180399 bp	DNA	linear	HTG 30-MAY-2000				
ACCESSION	Homo sapiens chromosome 11 clone RP11-686G6 map 11q22, WORKING								
VERSION	DRAFT SEQUENCE, 23 unordered pieces.								
KEYWORDS	AP001144								
SOURCE	AP001144.2 GI:8118496								
ORGANISM	HTG; HTGS_PHASE1; HTGS_DRAFT.								
	Homo sapiens (human)								
	Homo sapiens								
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;								
	Homnidae; Homo.								
REFERENCE	1 (bases 1 to 180399)								
AUTHORS	Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,								
	Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.								
TITLE	Homo sapiens 180,399 genomic DNA of 11q22								
JOURNAL	Published Only in Database (2000)								
REFERENCE	2 (bases 1 to 180399)								
AUTHORS	Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,								
	Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.								
TITLE	Direct Submission								
JOURNAL	Submitted (08-FEB-2000) Masahira Hattori, The Institute of Physical								
	and Chemical Research (RIKEN), Genomic Sciences Center (GSC),								
	Kitaado Univ., 1-15-1 Kitaado, Sagamihara, Kanagawa 228-8555,								
	Japan (E-mail:hattori@gsc.riken.go.jp,								
	URL:http://hbg.gsc.riken.go.jp/, Tel:81-42-778-9923,								
	Fax:81-42-778-9924)								
COMMENT	On May 31, 2000 this sequence version replaced gi:6937822.								
	----- Genome Center								
	Center: RIKEN Genomic Sciences Center (GSC)								
	Center code: RIKEN								
	Web site: http://hbg.gsc.riken.go.jp/								
	Contact: hattori@gsc.riken.go.jp								
	----- Project Information								
	Center project name: HumDrafc11								
	Center clone name: RP11-686G6								
	----- Summary Statistics								
	Sequencing vector: PCR products; 100% of reads								
	Chemistry: Dye-terminator ET-amersham; 100% of reads								
	Assembly program: Phrap; version 0.990129								
	Consensus quality: 163040 bases at least Q40								
	Consensus quality: 172128 bases at least Q30								
	Consensus quality: 176430 bases at least Q20								
	Insert size: 178199; sum-of-contigs								
	Quality coverage: 4.80x in Q20 bases; sum-of-contigs								

	NOTE: This is a 'working draft' sequence. It currently consists of								
	23 contigs. The true order of the pieces is not known and their								
	order in this sequence record is arbitrary. Gaps between the								
	contigs are represented as runs 'N', but the exact sizes of the gaps								
	are unknown. This record will be updated with the finished sequence								
	as soon as it is available and the accession number will be								
	preserved								
	1	32043 contig of	32043 bp in length						
	32144	52233 contig of	20090 bp in length						
	52334	70558 contig of	18225 bp in length						
	70659	83720 contig of	13065 bp in length						
	83821	91042 contig of	7222 bp in length						
	91143	100135 contig of	8993 bp in length						
	100236	108671 contig of</							

[illegible]


```

misc_feature      1..32043
                  /note="assembly_fragment"
misc_feature      32144..52233
                  /note="assembly_fragment clone_end:T7 vector_side:right"
misc_feature      52334..70558
                  /note="assembly_fragment"
misc_feature      70659..83720
                  /note="assembly_fragment"
misc_feature      83821..91042
                  /note="assembly_fragment"
misc_feature      91143..100135
                  /note="assembly_fragment"
misc_feature      100236..108671
                  /note="assembly_fragment"
misc_feature      108772..119459
                  /note="assembly_fragment"
misc_feature      119560..128905
                  /note="assembly_fragment"
misc_feature      129006..135545
                  /note="assembly_fragment"
misc_feature      135646..142043
                  /note="assembly_fragment"
misc_feature      142144..146798
                  /note="assembly_fragment"
misc_feature      146899..151826
                  /note="assembly_fragment"
misc_feature      151927..156030
                  /note="assembly_fragment"
misc_feature      156131..159677
                  /note="assembly_fragment"
misc_feature      159778..162873
                  /note="assembly_fragment"
misc_feature      162974..165817
                  /note="assembly_fragment"
misc_feature      165918..170216
                  /note="assembly_fragment"
misc_feature      170317..172721
                  /note="assembly_fragment"
misc_feature      172822..175136
                  /note="assembly_fragment"
misc_feature      175237..177561
                  /note="assembly_fragment"
misc_feature      177662..179149
                  /note="assembly_fragment"
misc_feature      179250..180399
                  /note="assembly_fragment"
misc_feature

```

ORIGIN

```

Query Match      42.8%; Score 27.4; DB 14; Length 180399;
Best Local Similarity 69.8%; Pred. No. 1.8e+02;
Matches 37; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

```

```

QY      3 TCCGCCGAGAAGCTTCAAAAGCGCTGAGAGAGCTGTTAGCAAAATCTGGAA 55
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      578 TCCGGGTAGAGGAGATTGTGCGCATGTAGATCTGATTAGAAATPAGCAA 526

```

RESULT 15

```

LOCUS      AX398707      34337 bp      DNA      linear      PAT 27-MAY-2002
DEFINITION Sequence 3 from Patent WO0220757.
ACCESSION  AX398707
VERSION     AX398707.1 GI:21261285
KEYWORDS

```

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.

REFERENCE

1. I. Z. C., Chaturvedi, K. C., Zhu, S. C., Woodage, T. C., Guegler, K. C.,
Webster, M. C., di Francesco, V. C. and Beasley, E. M.
Isolated human transporter proteins, nucleic acid molecules

JOURNAL encoding human transporter proteins, and uses thereof
Patent: WO 0220757-A 3 14-MAR-2002;
Applera Corporation (US)

FEATURES Location/Qualifiers
source 1..34337
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

```

Query Match      42.5%; Score 27.2; DB 6; Length 34337;
Best Local Similarity 72.9%; Pred. No. 2e+02;
Matches 35; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

```

```

QY      2 ATCCGCCGAGAAGCTTCAAAAGCGCTGAGAGAGCTGTTAGCAAAAT 49
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      21458 ATCCGCCGAGAAGCTTAAATGCCCGAGAGAGCTCTGGGCAATTAT 21505

```

Search completed: February 27, 2006, 11:11:20
JOD time : 373.475 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:36:56 ; Search time 361.222 Seconds
(without alignments)
8289.569 Million cell updates/sec

Title: US-10-789-164-6

Perfect score: 64
Sequence: 1 gatccgcgaagaagggtgtc.....aaatctggaactggaagta 64

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST.*

1: gb_esc1:*
2: gb_esc2:*
3: gb_esc3:*
4: gb_hic:*
5: gb_esc4:*
6: gb_esc5:*
7: gb_esc6:*
8: gb_esc7:*
9: gb_gsa1:*
10: gb_gsa2:*
11: gb_gsa3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	31.4	49.1	717	8	CX694562 ydb90d11.
2	29.4	45.9	588	9	BZ875603 CH240.194
3	29.2	45.6	248	9	AZ121609 RPT-33-4
4	28.6	44.7	299	1	AV178799 AV178799
5	28.6	44.7	300	1	AV182829 AV182829
6	28.6	44.7	300	1	AV200654 AV200654
7	28.6	44.7	300	5	C30991 C30991 Yuji
8	28.6	44.7	300	5	C54714 C54714 Yuji
9	28.6	44.7	300	5	C57993 C57993 Yuji
10	28.6	44.7	360	5	C63813 C63813 Yuji
11	28.6	44.7	385	8	D27678 CELK002D3R
12	28.6	44.7	608	3	Bj794144 Bj794144
13	28.6	44.7	667	3	Bj134236 Bj134236
14	28.6	44.7	669	3	Bj814155 Bj814155
15	28.6	44.7	677	3	Bj156049 Bj156049
16	28.6	44.7	703	3	Bj150848 Bj150848
17	28.6	44.7	726	3	Bj784546 Bj784546
18	28.6	44.7	749	3	Bj793333 Bj793333
19	28.6	44.7	762	3	Bj796838 Bj796838
20	28.4	44.4	1695	10	CL949298 OGI-FS004
21	28.4	44.1	300	5	CS1360 CS1360 Yuji
22	28.2	44.1	641	6	CD309732 StfPu691.

C 23	28.2	44.1	813	2	BF618952	BF618952 HVSMC000
C 24	28	43.8	536	10	CL368617	CL368617 RPT144_28
C 25	27.8	43.4	479	3	Bj159227	Bj159227 Bj159227
C 26	27.8	43.4	522	3	Bj160169	Bj160169 Bj160169
C 27	27.6	43.1	583	4	AY068267	AY068267 Schmidt
C 28	27.6	43.1	518	4	CN215701	CN215701 29524 Sub
C 29	27.6	43.1	901	3	BQ223859	BQ223859 AGENCOURT
C 30	27.2	42.5	571	3	ACQ79141	ACQ79141 RPT-11-2
C 31	27.2	42.5	614	8	DN428215	DN428215 LIB4216-1
C 32	27.2	42.5	1015	10	CD398108	CD398108 ZMBP0178
C 33	27	42.2	593	3	Bj558919	Bj558919 Bj558919
C 34	26.8	41.9	508	10	CL334959	CL334959 RPT144_25
C 35	26.6	41.6	374	8	N98037	N98037 2098C3 czap
C 36	26.6	41.6	426	5	BU495903	BU495903 PESTOab7
C 37	26.6	41.6	464	5	CA040335	CA040335 gsa1gcb51
C 38	26.6	41.6	467	6	CD882439	CD882439 Fl.106114
C 39	26.6	41.6	477	11	CR875824	CR875824 Sub scrof
C 40	26.6	41.6	500	1	AU087244	AU087244 AU087244
C 41	26.6	41.6	501	6	CB512806	CB512806 gsa1rgb54
C 42	26.6	41.6	508	8	CX357218	CX357218 gsa1rgb51
C 43	26.6	41.6	534	8	CX356249	CX356249 gsa1rgb51
C 44	26.6	41.6	538	10	CZ222011	CZ222011 A1RA-aae3
C 45	26.6	41.6	576	5	CA002465	CA002465 HS0712r

ALIGNMENTS

RESULT 1
CX694562 717 bp mRNA linear EST 19-JAN-2005
LOCUS ydb90d11.y2 Sea urchin EST Lib1. Strongylocentrotus purpuratus cDNA
DEFINITION Clone ydb90d11.5, mRNA sequence.

ACCESSION CX694562
VERSION CX694562.1 GI:57956947
KEYWORDS EST.

ORGANISM

Strongylocentrotus purpuratus
Strongylocentrotus purpuratus
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
Echinoidea; Euechinozoa; Echinozoa; Echinodermata; Echinozoa;
Strongylocentrotidae; Strongylocentrotus.

REFERENCE

Coffman, J.A., Robertson, A.J., Clifton, S., Pape, D., Hillier, L.,
Martin, J., Wyllie, T., Dante, M., Meyer, R., Theising, B., Bowers, Y.,
Gibbons, M., Konko, I., Teagareishvili, R., Rutter, B., Kennedy, S. and
Wilson, R.

Washu Sea Urchin EST Project
Unpublished (2004)

TITLE

Washu Sea Urchin EST Project
Contact: Dr. James A. Coffman
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800
Fax: 314 286 1810

Email: est@wustl.edu
DNA sequencing by: Washington University Genome Sequencing Center
Seq primer: -28RBPOT

High quality sequence stop: 717.
Location/Qualifiers

FEATURES

source
1. 717
/organism="Strongylocentrotus purpuratus"
/mol_type="mRNA"
/db_xref="taxon:7668"
/clone="ydb90d11"
/lab_host="DH108"
/clone_lib="Sea urchin EST Lib1"
/note="Vector: pCMVSPORT6.1 (Invitrogen); Site 1: NotI;
Site 2: SmaI. Arrayed normalized library of full-length
cDNAs representing blaetula stage transcriptome of the sea
urchin Strongylocentrotus purpuratus, cloned into the
vector pCMVSPORT6.1 (Invitrogen)"

ORIGIN

Query Match 49.1%; Score 31.4; DB 8; Length 717;
 Best Local Similarity 71.9%; Pred. No. 7;
 Matches 41; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

OY 6 GCGAAGAGGTTCACAAAGCCTGGAGAGCTGTTAGCAAAATCTGGAAGTGAAG 62
 DB 359 GCCAAGAGAGGTTCAGAGCAAAAGTGAAGCCGATAGATGAATGAGTAAGTGAAGTGAAG 415

RESULT 2
 BZ875603 588 bp DNA linear GSS 18-MAR-2003
 LOCUS CH240_194F12.TJ CHORI-240 Bos taurus genomic clone CH240_194F12,
 DEFINITION genomic survey sequence.
 ACCESSION BZ875603
 VERSION BZ875603.1 GI:29103008
 KEYWORDS GSS.
 SOURCE Bos taurus (cow)
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
 Pecora; Bovidae; Bovinae; Bos.
 1 (bases 1 to 588)
 Zhao,S., Shetty,J., Shatsman,S., Tsagaye,G., Geer,K.,
 Swartbeyn,A., Gebregeorgis,E., Chen,D., Riggs,F., de Jong,P.,
 Crawford,A.M. and McEwan,J.C.
 Bovine BAC End Sequences from Library CHORI-240
 Unpublished (2003)
 TITLE Contact: Shaying Zhao
 JOURNAL Department of Eukaryotic Genomics
 COMMENT The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the bovine BAC library CHORI-240
 (<http://www.chori.org/bacpac/bovine240.htm>). For BAC library
 availability, please contact Pieter de Jong (pjejong@mail.cno.org).
 Clones may be purchased from BACPAC Resources
 (http://www.chori.org/bacpac/ordering_information.htm). This work
 was undertaken as part of the International Bovine BAC Mapping
 Consortium (IBBMC) by AgResearch Ltd., New Zealand and the
 Institute of Genomic Research (TIGR), USA.
 Plate: 194 row: F column: 12
 Seq primer: SP6
 Class: BAC ends.

FEATURES
 source Location/Qualifiers
 1..588
 /organism="Bos taurus"
 /mol_type="genomic DNA"
 /strain="bred: Hereford"
 /db_xref="taxon:9913"
 /clone="CH240_194F12"
 /sex="Male"
 /cell_type="Blood"
 /clone_lib="CHORI-240"
 /note="Vector: pTARBAcl.3; Site.1: MboI; Site.2: MboI;
 Hereford bull li Domino 99375; CHORI-240 Bovine BAC
 library (Male) produced by Pieter de Jong"

ORIGIN
 Query Match 45.9%; Score 29.4; DB 9; Length 588;
 Best Local Similarity 70.9%; Pred. No. 32;
 Matches 39; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

OY 9 AAGAAGTGTTCACAAAGCCTGGAGAGCTGTTAGCAAAATCTGGAAGTGAAGT 63
 DB 140 ATGAAGAGGTGCAGAGACTGCACAGGCTCAGTCAAAAATCTGGAAGTGAAGT 194

RESULT 3
 A2121609/c 248 bp DNA linear GSS 12-MAY-2000
 LOCUS A2121609

DEFINITION RPCI-23-479G12.TJ RPCI-23 Mus musculus genomic clone
 RPCI-23-479G12, genomic survey sequence.
 ACCESSION A2121609
 VERSION A2121609.1 GI:7786685
 KEYWORDS GSS.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muridae; Muridae; Murinae; Mus.
 1 (bases 1 to 248)
 Zhao,S., Nierman,W., Feldlym,T., Malek,J., Shatsman,S.,
 Akniet,B., Levins,M., McGann,S., Tsagaye,G., Geer,K., Krol,M., de
 Jong,P. and Fraser,C.M.
 Mouse BAC End Sequences from Library RPCI-23
 Unpublished (1999)
 TITLE Other_GSSs: RPCI-23-479G12.TJ
 JOURNAL Contact: Shaying Zhao
 COMMENT Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-23. For BAC
 library availability, please contact Pieter de Jong
 (piet@dejong.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)
 or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 Plate: 479 row: G column: 12
 Seq primer: SP6
 Class: BAC ends.

FEATURES
 source Location/Qualifiers
 1..248
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-23-479G12"
 /sex="Female"
 /lab_host="DH10B"
 /clone_lib="RPCI-23"
 /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site.1:
 EcoRI; Site.2: EcoRI; Female C57BL/6J mouse kidney and/or
 brain genomic DNA was isolated and partially digested
 with a combination of EcoRI and EcoRI Methylase. Size
 selected DNA was cloned into the pBACe3.6 vector at the
 EcoRI sites. The ligation products were transformed into
 DH10B electrocompetent cells (BRL Life Technologies)."

ORIGIN
 Query Match 45.6%; Score 29.2; DB 9; Length 248;
 Best Local Similarity 69.0%; Pred. No. 32;
 Matches 40; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

OY 1 GATCGGCAAGAGGTTCACAAAGCCTGGAGAGCTGTTAGCAAAATCTGGAAGT 58
 DB 71 GAGCCAGAGAGAGGTTCAGATCCCGTGGAGCTGTGTAAGGCACTTGAAGT 14

RESULT 4
 AV178799/c 299 bp mRNA linear EST 22-NOV-2004
 LOCUS AV178799
 DEFINITION AV178799 Yui Kohara unpublished cDNA:Strain N2 hermaphrodite
 embryo Caenorhabditis elegans cDNA clone yk572b12.3', mRNA
 sequence.
 ACCESSION AV178799
 VERSION AV178799.2 GI:55937215
 KEYWORDS EST.
 SOURCE Caenorhabditis elegans
 ORGANISM Caenorhabditis elegans
 Eukaryota; Metazoa; Chordata; Chromadorea; Rhabdilita;

REFERENCE
AUTHORS

Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 299)
Kohara,Y., Shin-I,T., Thierry-Mieg,J., Thierry-Mieg,D., Mitsuiki,H.,
Nishigaki,A., Motohashi,T., Zeng,Q., Watanabe,H., Sugimoto,A.,
Sano,M., Miyata,A., Mitani,Y., Iida,K., Uesugi,H., Sugiyama,Y. and
Nomoto,H.

Expressed genes in C.elegans
Unpublished (1999)
On Jul 21, 1999 this sequence version replaced gi:5558700.
Contact: Yuji Kohara
Genome Biology Lab.

JOURNAL
COMMENT

National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
Location/Qualifiers

FEATURES
source

1. 299
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk572b12"
/sex="hermaphrodite"
/dev_stage="embryo"
/clone_1ib="Yuji Kohara unpublished cDNA:Strain N2
hermaphrodite embryo"

ORIGIN

Query Match 44.7%; Score 28.6; DB 1; Length 299;
Best Local Similarity 72.5%; Pred. No. 53;
Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

14 GGTGTTCAACGCCCTGAGAGCTGTTAGCAAAATCTGGAAGTGAAGTA 64
|||||
274 GGTGCTCAACCACTGTAGAGATGTGAGCAGCAATCTGTGTGGAAGA 224

RESULT 5
AV182829/c

300 bp mRNA linear EST 21-JUL-1999
AV182829 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite
embryo Caenorhabditis elegans cDNA clone yk643b6 3', mRNA sequence.

AV182829
AV182829.1 GI:5562730
EST.
Caenorhabditis elegans
Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.

REFERENCE
AUTHORS

1 (bases 1 to 300)
Kohara,Y., Shin-I,T., Thierry-Mieg,J., Thierry-Mieg,D., Mitsuiki,H.,
Nishigaki,A., Motohashi,T., Zeng,Q., Watanabe,H., Sugimoto,A.,
Sano,M., Miyata,A., Mitani,Y., Iida,K., Uesugi,H., Sugiyama,Y. and
Nomoto,H.

Expressed genes in C.elegans
Unpublished (1999)
Contact: Yuji Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
Location/Qualifiers

FEATURES
source

1. 300
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk643b6"
/sex="hermaphrodite"
/dev_stage="embryo"

ORIGIN

/clone_1ib="Yuji Kohara unpublished cDNA:Strain N2
hermaphrodite embryo"

Query Match 44.7%; Score 28.6; DB 1; Length 300;
Best Local Similarity 72.5%; Pred. No. 53;
Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

14 GGTGTTCAACGCCCTGAGAGCTGTTAGCAAAATCTGGAAGTGAAGTA 64
|||||
269 GGTGCTCAACCACTGTAGAGATGTGAGCAGCAATCTGTGTGGAAGA 219

Db

RESULT 6
AV200654/c 300 bp mRNA linear EST 26-JUL-1999
AV200654 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
clone yk578e6 3', mRNA sequence.

AV200654
AV200654.1 GI:5584425
EST.
Caenorhabditis elegans
Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.

REFERENCE
AUTHORS

1 (bases 1 to 300)
Kohara,Y., Shin-I,T., Thierry-Mieg,J., Thierry-Mieg,D., Mitsuiki,H.,
Nishigaki,A., Motohashi,T., Zeng,Q., Watanabe,H., Sugimoto,A.,
Sano,M., Miyata,A., Mitani,Y., Iida,K., Uesugi,H., Sugiyama,Y. and
Nomoto,H.

Expressed genes in C.elegans
Unpublished (1999)
Contact: Yuji Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
Location/Qualifiers

FEATURES
source

1. 300
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="CB1489 him-8(e1489)"
/db_xref="taxon:6239"
/clone="yk578e6"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"
/clone_1ib="Yuji Kohara unpublished cDNA"

ORIGIN

Query Match 44.7%; Score 28.6; DB 1; Length 300;
Best Local Similarity 72.5%; Pred. No. 53;
Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

14 GGTGTTCAACGCCCTGAGAGCTGTTAGCAAAATCTGGAAGTGAAGTA 64
|||||
170 GGTGCTCAACCACTGTAGAGATGTGAGCAGCAATCTGTGTGGAAGA 120

Db

RESULT 7
C30991 300 bp mRNA linear EST 18-OCT-1999
C30991 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite embryo
Caenorhabditis elegans cDNA clone yk267a12 3', mRNA sequence.

C30991
C30991.1 GI:2362787
EST.
Caenorhabditis elegans
Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.

REFERENCE 1 (bases 1 to 300)
AUTHORS Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
Sano,M., Miyata,A. and Nishigaki,A.
TITLE Expression map of the C.elegans genome
JOURNAL Unpublished (1996)
COMMENT Contact: Yuji Kohara
National Institute of Genetics
Genome Biology Lab.
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
Location/Qualifiers

FEATURES
source
1..300
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/clone="YK267a12"
/sex="hermaphrodite"
/dev_stage="embryo"
/clone_lib="Yuji Kohara unpublished cDNA:Strain N2
hermaphrodite embryo"

ORIGIN
Query Match 44.7%; Score 28.6; DB 5; Length 300;
Best Local Similarity 72.5%; Pred. No. 53; Mismatches 14; Indels 0; Gaps 0;
Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 14 GGTGTTCAACGCCCTGGAGAAGCTGTTAGCAAAATCTGGAACCTGGAAGTA 64
|||||
256 GGTGCTCAACCACTGTAAGATGTGGAGCAGAACTCCTGTGTGGAAGGA 206
|||||

RESULT 8
LOCUS C54714 300 bp mRNA linear EST 16-SEP-1997
DEFINITION C54714 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
clone yk397h12 3', mRNA sequence.
ACCESSION C54714
VERSION C54714.1 GI:2399315
KEYWORDS EST.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida;
Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 300)
AUTHORS Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
Sano,M., Miyata,A. and Nishigaki,A.
TITLE Expression map of the C.elegans genome
JOURNAL Unpublished (1996)
COMMENT Contact: Yuji Kohara
National Institute of Genetics
Genome Biology Lab.
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
Location/Qualifiers

FEATURES
source
1..300
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="CB1489 him-8(e1489)"
/db_xref="taxon:6239"
/clone="YK397h12"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"
/clone_lib="Yuji Kohara unpublished cDNA"

ORIGIN
Query Match 44.7%; Score 28.6; DB 5; Length 300;
Best Local Similarity 72.5%; Pred. No. 53; Mismatches 14; Indels 0; Gaps 0;
Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 14 GGTGTTCAACGCCCTGGAGAAGCTGTTAGCAAAATCTGGAACCTGGAAGTA 64
|||||
251 GGTGCTCAACCACTGTAAGATGTGGAGCAGAACTCCTGTGTGGAAGGA 201
|||||

RESULT 9
LOCUS C57993 300 bp mRNA linear EST 22-SEP-1997
DEFINITION C57993 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
clone yk311e7 3', mRNA sequence.
ACCESSION C57993
VERSION C57993.1 GI:2416698
KEYWORDS EST.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida;
Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 300)
AUTHORS Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
Sano,M., Miyata,A. and Nishigaki,A.
TITLE Expression map of the C.elegans genome
JOURNAL Unpublished (1996)
COMMENT Contact: Yuji Kohara
National Institute of Genetics
Genome Biology Lab.
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
Location/Qualifiers

FEATURES
source
1..300
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="CB1489 him-8(e1489)"
/db_xref="taxon:6239"
/clone="YK311e7"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"
/clone_lib="Yuji Kohara unpublished cDNA"

ORIGIN
Query Match 44.7%; Score 28.6; DB 5; Length 300;
Best Local Similarity 72.5%; Pred. No. 53; Mismatches 14; Indels 0; Gaps 0;
Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 14 GGTGTTCAACGCCCTGGAGAAGCTGTTAGCAAAATCTGGAACCTGGAAGTA 64
|||||
273 GGTGCTCAACCACTGTAAGATGTGGAGCAGAACTCCTGTGTGGAAGGA 223
|||||

RESULT 10
LOCUS C63813 360 bp mRNA linear EST 22-SEP-1997
DEFINITION C63813 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
clone yk351h10 5', mRNA sequence.
ACCESSION C63813
VERSION C63813.1 GI:2422518
KEYWORDS EST.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida;
Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 360)
AUTHORS Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
Sano,M., Miyata,A. and Nishigaki,A.
TITLE Expression map of the C.elegans genome
JOURNAL Unpublished (1996)
COMMENT Contact: Yuji Kohara
National Institute of Genetics

Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
Location/Qualifiers

FEATURES

source

1.360
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="CB1489 him-8(e1489)"
/db_xref="taxon:6239"
/clone="YK31h10"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"
/clone_1ib="Yuji Kohara unpublished cDNA"

ORIGIN

Query Match 44.7%; Score 28.6; DB 5; Length 360;
Best Local Similarity 72.5%; Pred. No. 55;
Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 14 GGTGTTCAACGCGCTGAGAGCTGTTTACCAAAATCTGGAACCTGGAAGTA 64
|||||
Db 293 GGTGCTCAACCACTGTAGAAATGTGAGCAGCAAAATCCTGTGTGGAAGGA 343

RESULT 11
D27678 385 bp mRNA linear EST 20-NOV-1995
LOCUS CELK00203R Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
DEFINITION clone YK2d3 3', mRNA sequence.
ACCESSION D27678
VERSION D27678.1 GI:522388
KEYWORDS EST.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 385)
AUTHORS Kohara, Y., Mitsuaki, H., Nishigaki, A., Motohashi, T., Sugimoto, A. and
Tabara, H.
Toward an expression map of the C.elegans genome
UNPUBLISHED (1994)
CONTACT: Yuji Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
Location/Qualifiers

FEATURES
source
1.385
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="CB1489 him-8(e1489)"
/db_xref="taxon:6239"
/clone="YK2d3"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"
/clone_1ib="Yuji Kohara unpublished cDNA"

ORIGIN

Query Match 44.7%; Score 28.6; DB 8; Length 385;
Best Local Similarity 72.5%; Pred. No. 56;
Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 14 GGTGTTCAACGCGCTGAGAGCTGTTTACCAAAATCTGGAACCTGGAAGTA 64
|||||
Db 319 GGTGCTCAACCACTGTAGAAATGTGAGCAGCAAAATCCTGTGTGGAAGGA 269

RESULT 12

BTJ94144/c 608 bp mRNA linear EST 25-MAY-2004
LOCUS BTJ94144 unpublished oligo-capped cDNA library Caenorhabditis
DEFINITION BTJ94144 clone YK1696f10 3', mRNA sequence.
ACCESSION BTJ94144
VERSION BTJ94144.1 GI:47672941
KEYWORDS EST.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 608)
AUTHORS Kohara, Y., Shin-I, T., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y.
and Sugano, S.
A complementary view of the C.elegans genome
UNPUBLISHED (2002)
CONTACT: Tadao Shin-I
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp.
Location/Qualifiers

FEATURES

source

1.608
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/clone="YK1696f10"
/sex="Hermaphrodite"
/tissue_type="whole animal"
/dev_stage="varied"
/clone_1ib="unpublished oligo-capped cDNA library"

ORIGIN

Query Match 44.7%; Score 28.6; DB 3; Length 608;
Best Local Similarity 72.5%; Pred. No. 61;
Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 14 GGTGTTCAACGCGCTGAGAGCTGTTTACCAAAATCTGGAACCTGGAAGTA 64
|||||
Db 314 GGTGCTCAACCACTGTAGAAATGTGAGCAGCAAAATCCTGTGTGGAAGGA 264

FEATURES

source

1.667
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/clone="YK1696f10"
/sex="Hermaphrodite"
/tissue_type="whole animal"
/dev_stage="varied"
/clone_1ib="unpublished oligo-capped cDNA library"

RESULT 13
BTJ34236 667 bp mRNA linear EST 23-JAN-2002
LOCUS BTJ34236 unpublished oligo-capped cDNA library, C. elegans L1 stage
DEFINITION Caenorhabditis elegans cDNA clone YK1086b09 3', mRNA sequence.
ACCESSION BTJ34236
VERSION BTJ34236.1 GI:18294393
KEYWORDS EST.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 667)
AUTHORS Kohara, Y., Shin-I, T., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y.
and Sugano, S.
A complementary view of the C.elegans genome
UNPUBLISHED (2002)
CONTACT: Tadao Shin-I
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp.
Location/Qualifiers

FEATURES

source

1.667
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/clone="YK1696f10"
/sex="Hermaphrodite"
/tissue_type="whole animal"
/dev_stage="varied"
/clone_1ib="unpublished oligo-capped cDNA library"

ORIGIN

Query Match 44.7%; Score 28.6; DB 8; Length 385;
Best Local Similarity 72.5%; Pred. No. 56;
Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 14 GGTGTTCAACGCGCTGAGAGCTGTTTACCAAAATCTGGAACCTGGAAGTA 64
|||||
Db 319 GGTGCTCAACCACTGTAGAAATGTGAGCAGCAAAATCCTGTGTGGAAGGA 269

/strain="N2"
/db_xref="taxon:6239"
/clone="yk1086b09"
/sex="hermaphrodite"
/tissue_type="whole animal"
/dev_stage="L1"
/clone_lib="unpublished oligo-capped cDNA library, C. elegans L1 stage"

ORIGIN

Query Match 44.7%; Score 28.6; DB 3; Length 667;
Best Local Similarity 72.5%; Pred. No. 63;
Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

OY 14 GGTTTCAAGCGCTGGAGAGCTGTTAGCAAAATCTGGAAGTGAAGTA 64
|||||
DB 515 GGTCCTCAACCACTGTAGAGATGTGGACAGAAATCCTGTGTGAAGGA 465
|||||

RESULT 14
Bj814155/c 669 bp mRNA linear EST 27-MAY-2004
DEFINITION Bj814155 unpublished oligo-capped cDNA library, stage L4
Caenorhabditis elegans cDNA clone yk1487d07 3', mRNA sequence.
Bj814155
ACCESSION Bj814155.1 GI:47723975
VERSION
KEYWORDS
SOURCE
ORGANISM
Caenorhabditis elegans
Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida;
Rhabdilitida; Rhabdilitidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 669)
Kohara, Y., Shin-I, T., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y.
and Sugano, S.
A complementary view of the C. elegans genome
Unpublished (2002)
Contact: Tadao Shin-I
National Institute of Genetics
Center For Genetic Resource Information
111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
Location/Qualifiers
1. 669
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk1487d07"
/sex="Hermaphrodite"
/tissue_type="whole animal"
/dev_stage="L4"
/clone_lib="unpublished oligo-capped cDNA library, stage L4"

ORIGIN

Query Match 44.7%; Score 28.6; DB 3; Length 669;
Best Local Similarity 72.5%; Pred. No. 63;
Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

OY 14 GGTTTCAAGCGCTGGAGAGCTGTTAGCAAAATCTGGAAGTGAAGTA 64
|||||
DB 370 GGTCCTCAACCACTGTAGAGATGTGGACAGAAATCCTGTGTGAAGGA 320
|||||

RESULT 15
Bj156049/c 677 bp mRNA linear EST 24-JAN-2002
LOCUS
DEFINITION Bj156049 unpublished oligo-capped cDNA library, C. elegans L1 stage
Caenorhabditis elegans cDNA clone yk1349c08 3', mRNA sequence.
Bj156049
ACCESSION
VERSION
Bj156049.1 GI:18324034

KEYWORDS
SOURCE
ORGANISM
EST.
Caenorhabditis elegans
Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida;
Rhabdilitida; Rhabdilitidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 677)
Kohara, Y., Shin-I, T., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y.
and Sugano, S.
A complementary view of the C. elegans genome
Unpublished (2002)
Contact: Tadao Shin-I
National Institute of Genetics
Center For Genetic Resource Information
111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
Location/Qualifiers
1. 677
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk1349c08"
/sex="hermaphrodite"
/tissue_type="whole animal"
/dev_stage="L1"
/clone_lib="unpublished oligo-capped cDNA library, C. elegans L1 stage"

ORIGIN

Query Match 44.7%; Score 28.6; DB 3; Length 677;
Best Local Similarity 72.5%; Pred. No. 63;
Matches 37; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

OY 14 GGTTTCAAGCGCTGGAGAGCTGTTAGCAAAATCTGGAAGTGAAGTA 64
|||||
DB 479 GGTCCTCAACCACTGTAGAGATGTGGACAGAAATCCTGTGTGAAGGA 429
|||||

Search completed: February 27, 2006, 11:09:37
Job time : 365.222 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocelebration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:34:16 ; Search time 40.4036 Seconds
(without alignments)
10556.990 Million cell updates/sec

Title: US-10-789-164-6

Sequence score: 64
1 gatccgcgaagaaggtgttc.....aaatctggaactggaagta 64

Scoring table: IDENTITY NUC
Gapop 10_0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21.*
1: geneseq1980s.*
2: geneseq1990s.*
3: geneseq2000s.*
4: geneseq2001as.*
5: geneseq2001bs.*
6: geneseq2002as.*
7: geneseq2002bs.*
8: geneseq2003as.*
9: geneseq2003bs.*
10: geneseq2003cs.*
11: geneseq2003ds.*
12: geneseq2004as.*
13: geneseq2004bs.*
14: geneseq2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	64	100.0	64	13	ADV65737	Adv65737 peptide A
2	60	93.8	64	13	ADV65738	Adv65738 Oligonuc
3	27.6	43.1	1706	6	ABL90642	Ab190642 Human pol
4	27.4	42.8	201	10	ADC98284	Adc98284 WMP1 gene
5	27.4	42.8	1257	10	ADT50834	Adt50834 Cancer re
6	27.4	42.8	81826	10	ADL13767	Adl13767 Osteocarc
7	27.4	42.8	81826	10	ADL13783	Adl13783 Osteocarc
8	27.4	42.8	81826	13	ADV35121	Adv35121 Human CDN
9	27.2	42.5	34337	6	ABL55857	Ab155857 Human GAB
10	27.2	42.5	34337	10	ADG88376	Adg88376 Human tra
11	26.8	41.9	78	12	ADL00404	Adl00404 Antipacte
12	26.4	41.2	1137	13	ADG51259	Adg51259 Bacteri
13	26.4	41.2	1578	11	ADM03140	Adm03140 Human CDN
14	26	40.6	2681	2	AAT68323	Aat68323 Clarkia b
15	26	40.6	2681	2	AAV08878	Aav08878 Linalool
16	26	40.6	2681	3	AAZ52308	Aaz52308 Clarkia b
17	26	40.6	2760	10	AAAD47173	Aad47173 C. brewer
18	26	40.6	3708	10	ACF57816	Acf57816 C. brewer
19	26	40.6	4434	9	ADB10028	Adb10028 Allotiococ

20	26	40.6	4434	9	ADB10026	Adb10026 Allotiococ
21	26	40.6	110000	9	ADB12064_09	Continuation (10 o
22	25.6	40.0	350	4	AAK56389	Aak56389 Human imm
23	25.6	40.0	426	13	ACF90670	Acf90670 Human SIR
24	25.6	40.0	1650	12	ACH87010	Ach87010 Human gen
25	25.6	40.0	2350	4	AAH45798	Aah45798 Human cel
26	25.6	40.0	3559	6	AB199910	Ab199910 Human sec
27	25.6	40.0	3732	6	AB199842	Ab199842 Human sec
28	25.6	40.0	10953	5	AAS29204	Aas29204 Genomic s
29	25.6	40.0	10953	6	ABS68344	Abs68344 Human DNA
30	25.6	40.0	10953	10	ADC25466	Adc25466 Human CDN
31	25.6	40.0	98865	6	ABQ78054	Abq78054 Human Ras
32	25.6	40.0	110000	6	ABQ69245_13	Continuation (14 o
33	25.6	40.0	110000	6	ABQ67195_3	Continuation (4 of
34	25.4	39.7	834	12	ADL03259	Adl03259 DNA encod
35	25.4	39.7	269223	4	AAF28554	Aaf28554 Genomic f
36	25.2	39.4	311	4	ABW71845	Abw71845 Human foe
37	25.2	39.4	311	4	AA152189	Aa152189 Probe #20
38	25.2	39.4	311	4	AAK46290	Aak46290 Human bon
39	25.2	39.4	311	4	AAK20221	Aak20221 Human bra
40	25.2	39.4	311	4	ABS46013	Abs46013 Human liv
41	25.2	39.4	311	6	ABS20604	Abes20604 Human gen
42	25.2	39.4	443	4	ABW59333	Abw59333 Human foe
43	25.2	39.4	443	4	AA139134	Aa139134 Probe #78
44	25.2	39.4	443	4	AAK33347	Aak33347 Human bon
45	25.2	39.4	443	4	AAK07551	Aak07551 Human bra

ALIGNMENTS

RESULT 1
ADV65737
ID ADV65737 standard; DNA, 64 BP.

XX ADV65737;

DT 10-FEB-2005 (first entry)

XX Peptide Anal3 encoding DNA SEQ ID NO:6.

XX ds; gene expression; antibiotic; antimicrobial; fungicide; cytostatic;
XX peptide Anal3.

XX Synthetic.

XX KR2004034780-A.

XX 29-APR-2004.

XX 17-OCT-2002; 2002KR-00063379.

XX 17-OCT-2002; 2002KR-00063379.

XX (BIOL-) BIOLEADERS CORP.

XX (UYCH-) UNIV CHUNGSUN CO LTD.

XX (KOR-) KOREA RES INST BIOTSCIENCE & BIOTECHNOLOG.

XX Boe HR, Ham GS, Hong SP, Jung CM, Kim CJ, Lee DG, Lee JS;

XX Park YG, Sung MH;

XX WPI, 2004-577380/56.

XX Method for surface expression of peptides p5 and anal3 using pgs bca

XX gene.

XX Example 3; SEQ ID NO 6; 25pp; Korean.

XX The invention relates to a novel method for surface expression of

XX peptides p5 and Anal3 using a poly-gamma-glutamic acid synthase (pgs)BCA

XX gene, thereby removing a purification process of peptides p5 and Anal3,

XX and using lactic acid bacteria for the surface expression, so that

XX peptide antibiotics can be cheaply and stably mass-produced. An

CC expression vector pHELIB:pgsa-P5 comprises one or more genes encoding
CC poly-gamma-glutamic acid synthase selected from pgsB, pgsC and pgsA, and
CC a gene for dipolar peptide antibiotics having antimicrobial, antifungal
CC and anticancer activities, wherein the dipolar peptide antibiotic has
CC homology to the peptide P5 encoded by the nucleotide sequence set forth
CC in ADV65735, or to the peptide Anal3 encoded by the nucleotide sequence
CC set forth in ADV65737. The present sequence encodes the peptide Anal3
CC used in the invention.
XX
SQ Sequence 64 BP; 21 A; 11 C; 19 G; 13 T; 0 U; 0 Other;
Query Match 100.0%; Score 64; DB 13; Length 64;
Best Local Similarity 100.0%; Pred. No. 6.5e-13;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 GATCGCGAAGAGGTGTTCAACGCCCTGGAGAACTGTTTAGCAAAATCTGGAACCTGGA 60
DB 1 GATCGCGAAGAGGTGTTCAACGCCCTGGAGAACTGTTTAGCAAAATCTGGAACCTGGA 60
OY 61 AGTA 64
DB 61 AGTA 64
Db 61 AGTA 64
RESULT 2
ADV65738/C
ID ADV65738 standard; DNA; 64 BP.
XX
AC ADV65738;
XX
DT 10-FEB-2005 (first entry)
XX
DE Oligonucleotide of the invention SEQ ID NO:7.
XX
KW ss; gene expression; antibiotic; antimicrobial; fungicide; cytostatic.
XX
OS Synthetic.
XX
PN KR2004034780-A.
XX
PD 29-APR-2004.
XX
PF 17-OCT-2002; 2002KR-00063379.
XX
PR 17-OCT-2002; 2002KR-00063379.
XX
PA (BIOL-) BIOLEADERS CORP.
XX
PA (UYCH-) UNIV CHUNSUN CO LTD.
XX
PA (KORE-) KOREA RES INST BIOSCIENCE & BIOTECHNOLOG.
XX
PI BOO HR, Ham GS, Hong SP, Jung CM, Kim CJ, Lee DG, Lee JS;
PI Park YG, Sung MH;
XX
DR WPI; 2004-577380/56.
XX
PT Method for surface expression of peptides p5 and anal3 using pgs bca
PT gene.
XX
PS Example 3; SEQ ID NO 7; 25bp; Korean.
XX
CC The invention relates to a novel method for surface expression of
CC peptides P5 and Anal3 using a poly-gamma-glutamic acid synthase (pgs) bca
CC gene, thereby removing a purification process of peptides P5 and Anal3,
CC and using lactic acid bacteria for the surface expression, so that
CC peptide antibiotics can be cheaply and stably mass-produced. An
CC expression vector pHELIB:pgsa-P5 comprises one or more genes encoding
CC poly-gamma-glutamic acid synthase selected from pgsB, pgsC and pgsA, and
CC a gene for dipolar peptide antibiotics having antimicrobial, antifungal
CC and anticancer activities, wherein the dipolar peptide antibiotic has
CC homology to the peptide P5 encoded by the nucleotide sequence set forth
CC in ADV65735, or to the peptide Anal3 encoded by the nucleotide sequence
CC set forth in ADV65737. The present sequence represents an oligonucleotide
CC used in the invention to introduce the peptide Anal3 encoding DNA into

CC expression vector pHELIB:A-Anal3.
XX
SQ Sequence 64 BP; 14 A; 19 C; 11 G; 20 T; 0 U; 0 Other;
Query Match 93.8%; Score 60; DB 13; Length 64;
Best Local Similarity 100.0%; Pred. No. 1.7e-11;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 5 CGCGAAGAGGTGTTCAACGCCCTGGAGAACTGTTTAGCAAAATCTGGAACCTGGA 64
DB 64 CGCGAAGAGGTGTTCAACGCCCTGGAGAACTGTTTAGCAAAATCTGGAACCTGGA 5
RESULT 3
ABL90642
ID ABL90642 standard; cDNA; 1706 BP.
XX
AC ABL90642;
XX
DT 24-MAY-2002 (first entry)
XX
DE Human polynucleotide SEQ ID NO 1204.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
KW vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein; gene; ss.
XX
XX Homo sapiens.
XX
PN WO200190304-A2.
XX
PD 29-NOV-2001.
XX
PF 18-MAY-2001; 2001WO-US016450.
XX
PR 19-MAY-2000; 2000US-0205515P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Birse CE, Rosen CA;
XX
DR WPI; 2002-122018/16.
XX
DR P-PSDB; ABB90233.
XX
PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
PT prevention of neural, immune system, muscular, reproductive,
PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
PT disorders.
XX
PS Claim 4; SEQ ID NO 1204; 2081bp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABL89449-ABL90853) and proteins
CC (ABB89040-ABB90444) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing
CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC infections. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pce_sequences
XX
SQ Sequence 1706 BP; 527 A; 350 C; 415 G; 410 T; 0 U; 4 Other;

Query Match 43.1%; Score 27.6; DB 6; Length 1706;
Best Local Similarity 67.2%; Pred. No. 14;
Matches 39; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 4 CCGGCAAGAGGTGTTCAACGCCCTGAGAGAGCTGTTAGCAAAATCTGAA 61
DB 57 CCATGAAGAGGATTCAACCGCATGAAAGCAGCTGCTTACCAAGACCTGGGCAAGAA 114

RESULT 4

AD98284
ID ADC98284 standard; DNA; 201 BP.

AC ADC98284;
XX

DT 01-JAN-2004 (first entry)
XX

DE MMP1 gene related polymorphism marker nucleotide sequence.
XX

XX low bone mineral density; BMD; bone damage; polymorphism; osteoporosis;
KM single nucleotide polymorphism; SNP; gene; ds; human.
XX

OS Homo sapiens.
XX

PN WO2003054218-A2.
XX

PD 03-JUL-2003.
XX

PF 19-DEC-2002; 2002WO-US040948.
XX

PR 20-DEC-2001; 2001US-0342711P.
XX

PR 04-NOV-2002; 2002US-0423559P.
XX

XX (INCYTE GENOMICS INC.
PA

PI Jones KA, Valdes A, Townley DJ, Mangion J, Galwey N, Bennett S;
PI McKay I, Schacter A;
XX

DR WPI; 2003-559156/52.
XX

PT Determining whether an individual is predisposed to susceptibility to low
bone mineral density (BMD) and/or bone damage, involves identifying
polymorphisms in associated genes.
XX

PS Claim 4; Page 215; 246pp; English.
XX

CC The present invention describes a method of determining whether an
individual is predisposed to susceptibility to low bone mineral density
(BMD) and/or bone damage comprising identifying whether the individual
has at least one polymorphism in a polynucleotide encoding a protein,
CC where the polynucleotide is one of 81 200-500 nucleotide sequences (S1,
see ADC98235 to ADC98315). An agent identified in an method from the
CC present invention which can be used for the prevention or treatment of a
disease resulting in susceptibility to low BMD and/or bone damage is
CC useful in the manufacture of a medicament for use in modulating the
CC susceptibility to low BMD and/or bone damage. The disease associated with
CC low BMD and/or bone damage is osteoporosis. The present sequence is used
in the exemplification of the present invention.
XX

XX Sequence 201 BP; 62 A; 38 C; 46 G; 54 T; 0 U; 1 Other;

Query Match 42.8%; Score 27.4; DB 10; Length 201;
Best Local Similarity 69.8%; Pred. No. 9.3;
Matches 37; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 3 TCCGCGAAGAGGTGTTCAACGCCCTGAGAGAGCTGTTAGCAAAATCTGAA 55
DB 34 TCCGCGTGAAGAGGATTGTGCGCATGTAGATCTGATTGAAATAAGCAA 86

RESULT 5
ADT50834/C
ID ADT50834 standard; DNA; 1257 BP.

XX ADT50834;
AC
XX 13-JAN-2005 (first entry)
DT
XX
XX
DE
XX
XX

XX Cancer related nucleic acid sequence #138.

KM ds; gene; cytostatic; gene therapy; vaccine; diagnosis; breast; colon;
lung; ovarian; prostate; cancer.
XX

OS Homo sapiens.
XX

PN WO2004092338-A2.
XX

PD 28-OCT-2004.
XX

PF 12-APR-2004; 2004WO-US011104.
XX

PR 11-APR-2003; 2003US-0462399P.
XX

PR 01-JUL-2003; 2003US-0484333P.
XX

PA (DIAD-) DIADEXUS INC.
XX

PI Macina RA, Turner LR, Sun Y, Tam A;
XX

DR WPI; 2004-766851/75.
XX

PT New cancer specific nucleic acid (CasNA) molecules, useful for
diagnosing, monitoring the presence of, or treating a patient with
breast, colon, lung, ovarian, or prostate cancer.
XX

PS Claim 1, SEQ ID NO 138; 891pp; English.
XX

CC The invention relates to an isolated nucleic acid molecule (I)
selectively hybridizing to, or comprising at least 95% sequence identity
to, any of the 362 nucleotide sequences fully defined in the
CC specification. The nucleic acid molecules and polypeptides are useful for
diagnosing, monitoring the presence of, or treating a patient with
breast, colon, lung, ovarian, or prostate cancer. This sequence
CC corresponds to a nucleic acid of the invention.
XX

XX Sequence 1257 BP; 397 A; 240 C; 245 G; 375 T; 0 U; 0 Other;

Query Match 42.8%; Score 27.4; DB 13; Length 1257;
Best Local Similarity 69.8%; Pred. No. 15;
Matches 37; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 3 TCCGCGAAGAGGTGTTCAACGCCCTGAGAGAGCTGTTAGCAAAATCTGAA 55
DB 275 TCCGCGTGAAGAGGATTGTGCGCATGTAGATCTGATTGAAATAAGCAA 223

RESULT 6

ADL13767
ID ADL13767 standard; DNA; 81826 BP.

AC ADL13767;
XX

DT 06-MAY-2004 (first entry)
XX

DE Osteoarthritis-associated polymorphic nucleotide #299.
XX

XX ds; gene; osteopathic; antiinflammatory; antiarthritic; gene therapy;
KM joint space narrowing; osteophyte development; joint pain;
XX osteoarthritis; SNP; single nucleotide polymorphism.
XX

OS Homo sapiens.
XX

PN WO2003054166-A2.
XX

PD 03-JUL-2003.
XX

PF 19-DEC-2002; 2002WO-US041225.
XX

```
XX 20-DEC-2001; 2001US-0342603P.
XX (INCY-) INCYTE GENOMICS INC.
XX Jones KA, Schafer A;
XX WPI; 2003-559141/52.
XX
XX Determining susceptibility of an individual to joint space narrowing,
XX osteophyte development and/or joint pain comprises identifying whether
XX the individual has at least one polymorphism in a polynucleotide encoding
XX a protein.
XX
XX Disclosure: SEQ ID NO 299; 297bp; English.
XX
XX The invention relates to a method of determining susceptibility of an
XX individual to joint space narrowing and/or osteophyte development and/or
XX joint pain comprising identifying whether the individual has at least one
XX polymorphism in a polynucleotide encoding at least one of the protein
XX listed in the specification. The methods, composition and agent are
XX useful for modulating the susceptibility of an individual to joint space
XX narrowing and/or osteophyte development and/or joint pain that is
XX associated with a disease, preferably osteoarthritis. The cell line and
XX the non-human animal are useful for screening for an agent for diagnosing
XX an individual having susceptibility to joint space narrowing and/or
XX osteophyte development and/or joint pain. This sequence corresponds to
XX the polynucleotide encoding a protein listed in the specification. (Note:
XX The sequence data for this patent did not form part of the printed
XX specification but was obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences).
XX
XX Sequence 81826 BP; 23858 A; 16501 C; 15287 G; 26165 T; 0 U; 15 Other;
XX
XX Query Match 42.8%; Score 27.4; DB 10; Length 81826;
XX Best Local Similarity 69.8%; Pred. No. 45;
XX Matches 37; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
XX
XX 3 TCCGCGAAGAGGTGTCGAACGCTGGAGAACTGTTAGCAAAATCTGGA 55
XX 6519 TCCGGGTGAAAGGATTGTGCGCATGTAGATCTGATTAGAAATAAGCAA 6571
XX
XX RESULT 7
XX ADL13783
XX ID ADL13783 standard; DNA; 81826 BP.
XX AC ADL13783;
XX
XX 06-MAY-2004 (first entry)
XX
XX Osteoarthritis-associated polymorphic nucleotide #315.
XX
XX ds; gene; osteopathic; antiinflammatory; antiarthritic; gene therapy;
XX joint space narrowing; osteophyte development; joint pain;
XX osteoarthritis; SNP; single nucleotide polymorphism.
XX
XX Homo sapiens.
XX
XX WO2003054166-A2.
XX
XX 03-JUL-2003.
XX
XX 19-DEC-2002; 2002MO-US041225.
XX
XX 20-DEC-2001; 2001US-0342603P.
XX (INCY-) INCYTE GENOMICS INC.
XX Jones KA, Schafer A;
XX WPI; 2003-559141/52.
XX
```

```
PT Determining susceptibility of an individual to joint space narrowing,
PT osteophyte development and/or joint pain comprises identifying whether
PT the individual has at least one polymorphism in a polynucleotide encoding
PT a protein.
PT
PT Disclosure: SEQ ID NO 315; 297bp; English.
PT
PT The invention relates to a method of determining susceptibility of an
PT individual to joint space narrowing and/or osteophyte development and/or
PT joint pain comprising identifying whether the individual has at least one
PT polymorphism in a polynucleotide encoding at least one of the protein
PT listed in the specification. The methods, composition and agent are
PT useful for modulating the susceptibility of an individual to joint space
PT narrowing and/or osteophyte development and/or joint pain that is
PT associated with a disease, preferably osteoarthritis. The cell line and
PT the non-human animal are useful for screening for an agent for diagnosing
PT an individual having susceptibility to joint space narrowing and/or
PT osteophyte development and/or joint pain. This sequence corresponds to
PT the polynucleotide encoding a protein listed in the specification. (Note:
PT The sequence data for this patent did not form part of the printed
PT specification but was obtained in electronic format directly from WIPO at
PT ftp.wipo.int/pub/published_pct_sequences).
PT
PT Sequence 81826 BP; 23858 A; 16501 C; 15287 G; 26165 T; 0 U; 15 Other;
PT
PT Query Match 42.8%; Score 27.4; DB 10; Length 81826;
PT Best Local Similarity 69.8%; Pred. No. 45;
PT Matches 37; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
PT
PT 3 TCCGCGAAGAGGTGTCGAACGCTGGAGAACTGTTAGCAAAATCTGGA 55
PT 6519 TCCGGGTGAAAGGATTGTGCGCATGTAGATCTGATTAGAAATAAGCAA 6571
PT
PT RESULT 8
PT ADV35121
PT ID ADV35121 standard; CDNA; 81826 BP.
PT AC ADV35121;
PT
PT 10-FEB-2005 (first entry)
PT
PT Human CDNA of an exemplary efficacy gene for BAD SegID197.
PT
PT human; ss; multi-parameter high throughput screening; MPHTS;
PT disease signature; neuropsychiatric; neurodegenerative; schizophrenia;
PT bipolar affective disorder; BAD; autism; Parkinson's;
PT Alzheimer's disease; neuroleptic; nootropic; anti manic; antidepressant.
PT
PT Homo sapiens.
PT
PT US2003096264-A1.
PT
PT 22-MAY-2003.
PT
PT 18-JUN-2002; 2002US-00175523.
PT
PT 18-JUN-2001; 2001US-0299151P.
PT 07-SEP-2001; 2001US-0317828P.
PT 25-SEP-2001; 2001US-0325150P.
PT 14-NOV-2001; 2001US-033047P.
PT 18-JAN-2002; 2002US-0349936P.
PT 04-MAR-2002; 2002US-0361834P.
PT
PT (PSYC-) PSYCHIATRIC GENOMICS INC.
PT Altar CA, Brockman JA, Evans D, Hook D, Klimczak LJ, Iaeng P;
PT Palfreyman M, Rajan P;
PT WPI; 2004-118903/12.
PT
PT Identifying a compound that can treat disease or disorders, such as, a
PT neuropsychiatric disorder e.g., schizophrenia, or autism, comprises
```

PT determining the expression of one or more efficacy genes in a cell
PT contacted with the test compound.

Example 6; SEQ ID NO 197; 39pp; English.

CC This invention relates to a novel screening method identified as a multi-
CC parameter high throughput screening (MPHS) assay. Specifically, it
CC refers to an assay that utilizes the disease signature of a plurality of
CC specific genes associated with a particular disease, and identifies
CC differential expression between those cells taken from individuals
CC affected by that disease and those that are not affected. The present
CC invention then describes the screening of candidate pharmaceutical
CC compounds to identify those that have a potential therapeutic benefit for
CC the treatment of neuropsychiatric and neurodegenerative disorders
CC including schizophrenia, bipolar affective disorder (BAD) and autism, as
CC well as Parkinson's and Alzheimer's disease. Accordingly, the compounds
CC of this invention exhibit various activities including neuroleptic,
CC neurotropic, anti manic and antidepressant. Furthermore, the screening
CC method used in MPHS will be automated, such that a large number of test
CC compounds may be rapidly screened with a minimal amount of labour and
CC effort. This polynucleotide is a human cDNA sequence of a gene that is
CC differentially expressed in the presence of a therapeutic compound and
CC represents an exemplary efficacy gene for bipolar affective disorder,
CC given in an exemplification of the invention.

XX Sequence 81826 BP; 23858 A; 16501 C; 15287 G; 26165 T; 0 U; 15 Other;

Query Match 42.8%; Score 27.4; DB 13; Length 81826;

Best Local Similarity 69.8%; Pred. No. 45;
Matches 37; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 3 TCCGCGAAGAGGTGTTCAACGCTTGAGACCTGTTGCAAAATCTGGA 55
Db 6519 TCCGGGTAGAAAGGATTTGTGCGCATGTAGAAATCTGATTAAGAAATAGCAA 6511

RESULT 9
ABLS5857
ID ABL55857 standard; DNA; 34337 BP.

XX ABL55857;

XX 15-JUL-2002 (first entry)

XX Human GABA transporter protein gene.

XX Human; GABA; transporter protein; drug screening; therapeutic target;

XX gene; ds.

XX Homo sapiens.

XX Key Location/Qualifiers

XX variation replace(701,C/T)
/*tag= a
/standard_name= "Single nucleotide polymorphism"

XX CDS 2001. .32337
/*tag= b
/product= "GABA transporter protein"

XX exon /note= "Contains 10 introns"

XX 2001. .2289
/*tag= c
/number= 1

XX intron 2290. .8039
/*tag= d
/number= 1

XX variation replace(4738,C)
/*tag= e
/standard_name= "Single nucleotide polymorphism"

XX variation replace(6156,A/G/C)
/*tag= f
/standard_name= "Single nucleotide polymorphism"

XX variation replace(6425,C)
/*tag= g

XX variation

XX variation

FT /standard_name= "Single nucleotide polymorphism"
FT replace(6444,G)
/*tag= h

FT /standard_name= "Single nucleotide polymorphism"
FT replace(7538,A)
/*tag= i

FT /standard_name= "Single nucleotide polymorphism"
FT 8040. .8197
/*tag= j

FT intron /number= 2
8198. .8564
/*tag= k

FT exon /number= 2
8547. .8673
/*tag= l

FT intron /number= 3
8674. .10067
/*tag= m

FT variation replace(9127,G)
/*tag= n

FT variation replace(9256,G)
/*tag= o

FT variation /standard_name= "Single nucleotide polymorphism"
FT replace(9421,G)
/*tag= p

FT exon /standard_name= "Single nucleotide polymorphism"
FT 10068. .10249
/*tag= q

FT intron /number= 4
10250. .17500
/*tag= r

FT variation /number= 4
13248. .13250,TC)
/*tag= s

FT variation replace(13377,G)
/*tag= t

FT variation replace(13434,C)
/*tag= u

FT variation /standard_name= "Single nucleotide polymorphism"
FT replace(13672,A)
/*tag= v

FT variation /standard_name= "Single nucleotide polymorphism"
FT replace(14502,T)
/*tag= w

FT variation /standard_name= "Single nucleotide polymorphism"
FT replace(14774. .14776,AAA/AA)
/*tag= x

FT variation /standard_name= "Single nucleotide polymorphism"
FT replace(17289,A)
/*tag= y

FT exon /standard_name= "Single nucleotide polymorphism"
FT 17501. .17611
/*tag= z

FT intron /number= 5
17612. .20639
/*tag= aa

FT variation replace(17660,G)
/*tag= ab

FT variation /standard_name= "Single nucleotide polymorphism"
FT replace(19702,G)
/*tag= ac

FT exon /standard_name= "Single nucleotide polymorphism"
FT 20640. .20881
/*tag= ad

FT intron /number= 6
20882. .21173
/*tag= ae

FT /number= 6

FT exon 21174..21366
FT /*tag= af
FT /number= 7
FT replace(21263,T)
FT /*tag= ag
FT /standard_name= "single nucleotide polymorphism"
FT 21367..21397
FT /*tag= ah
FT /number= 7
FT 23298..23490
FT /*tag= ai
FT /number= 8
FT 23491..26774
FT /*tag= aj
FT /number= 8
FT replace(24502..24504,TC)
FT /*tag= ak
FT /standard_name= "single nucleotide polymorphism"
FT replace(24642,C)
FT /*tag= al
FT /standard_name= "single nucleotide polymorphism"
FT 26775..26934
FT /*tag= am
FT /number= 9
FT 26935..30367
FT /*tag= an
FT /number= 9
FT replace(28543,T)
FT /*tag= ao
FT /standard_name= "single nucleotide polymorphism"
FT replace(30303,C)
FT /*tag= ap
FT /standard_name= "single nucleotide polymorphism"
FT 30368..30530
FT /*tag= aq
FT /number= 10
FT 30531..31962
FT /*tag= ar
FT /number= 10
FT 31963..32337
FT /*tag= as
FT /number= 11
FT exon
FT US2002031800-A1.
FT 14-MAR-2002.
FT 21-DEC-2000; 2000US-00741149.
FT 05-SEP-2000; 2000US-0229529P.
FT exon
FT (LITZ/) LI Z.
FT (CHAT/) CHATURVEDI K.
FT (ZHUS/) ZHU S.
FT (WOOD/) WOODAGE T.
FT (GUEG/) GUEGLER K.
FT (WEBER/) WEBSTER M.
FT (DFRA/) DI FRANCESCO V.
FT (BEAS/) BEASLEY E M.
FT Li Z, Chaturvedi K, Zhu S, Woodage T, Guegler K, Webster M;
FT Di Francesco V, Beasley EM;
FT WPI; 2002-361179/39.
FT P-PSDB; ABB71168.
FT exon
FT New peptides and nucleic acid sequences related to the GABA transporter
FT subfamily, useful in developing diagnostic compositions, as well as in
FT drug screening, particularly as models for the development of human
FT therapeutic targets.
FT Claim 2; Fig 3; 69pp; English.
FT ,XX

CC The sequence encodes a novel human transporter protein of the GABA
CC transporter subfamily. The invention relates to a novel isolated human
CC peptide of the GABA transporter subfamily. The peptide and nucleic acids
CC are useful in developing human therapeutics and diagnostic compositions.
CC These are also useful in drug screening, particularly as models for the
CC development of human therapeutic targets
XX
SQ Sequence 34337 BP; 10987 A; 5399 C; 6061 G; 11588 T; 0 U; 302 Other;
Query Match 42.5%; Score 27.2; DB 6; Length 34337;
Best Local Similarity 72.9%; Pred. No. 42;
Matches 35; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
OY 2 ATCCGCGAAGAGGTGTTCAACCGCTGAGAGCTGTTAGCAAAAT 49
DB 21458 ATGCCCGAAGAGCTCTTAATAATGCCGGAAGAGCTCCTGGCATTAAT 21505
RESULT 10
ADG88376
ID ADG88376 standard; DNA; 34337 BP.
XX
AC ADG88376;
XX
DT 11-MAR-2004 (first entry)
XX
DE Human transporter gene.
XX
KW Human; transporter protein; GABA transporter; therapeutic protein;
KW pharmacogenomic analysis; immune response; biological fluid;
KW foetal heart tissue; gene therapy; gene; db.
XX
OS Homo sapiens.
XX
FH Key
FH CDS
FT Location/Qualifiers
FT 1001..31067
FT /*tag= b
FT /product= "Human transporter protein"
FT 1001..1289
FT /*tag= a
FT 1290..7053
FT /*tag= c
FT 7054..7211
FT /*tag= d
FT 7212..7560
FT /*tag= e
FT 7561..7687
FT /*tag= f
FT 7688..9081
FT /*tag= g
FT 9082..9263
FT /*tag= h
FT 9264..16514
FT /*tag= i
FT 16515..16625
FT /*tag= j
FT 16626..19366
FT /*tag= k
FT 19367..19608
FT /*tag= l
FT 19609..19900
FT /*tag= m
FT 19901..20093
FT /*tag= n
FT 20094..22026
FT /*tag= o
FT 22027..22219
FT /*tag= p
FT 22220..25504
FT /*tag= q
FT 25505..25664
FT /*tag= r
FT 25665..29097
FT intron

FT FT /tag= 8
FT exon 29098. 29260
FT /tag= t
FT intron 29261. 30692
FT /tag= u
FT exon 30693. 31067
FT /tag= v
XX US2003157649-A1.
XX 21-AUG-2003.
XX 12-MAR-2003; 2003US-00385614.
XX 05-SEP-2000; 2000US-0229529P.
XX 21-DEC-2000; 2000US-00741149.
XX 05-SEP-2001; 2001WO-US027403.
XX (APPL-) APPLERA CORP.
XX Li Z, Chaturvedi K, Zhu S, Woodage T, Guejler K, Webster M;
XX Di Francesco V, Beasley EM;
XX WPI: 2003-777995/73.
XX P-PSDB; ADG88375.
XX DR
XX New human transporter proteins and nucleic acid molecules, useful for
XX PT diagnosing, preventing or treating disorders associated with aberrant
XX PT expression of the transporter proteins, or in pharmacogenomic analysis.
XX
XX Claim 4; SEQ ID NO 3; 59pp; English.
XX
XX The invention relates to human transporter protein and its corresponding
XX CC invention acid sequence related to GABA transporter subfamily. The
XX CC invention is useful as models for the development of human therapeutic
XX CC targets, aids in the identification of therapeutic proteins, serves as
XX CC targets for the development of human therapeutic agents that modulate
XX CC transporter activity in cells and tissues that express the transporter,
XX CC or in pharmacogenomic analysis. The transporter peptide may be used in
XX CC substantial and specific assays, in raising antibodies or in eliciting
XX CC another immune response, as a reagent in assays designed to
XX CC quantitatively determine the levels of the protein in biological fluids,
XX CC and as markers for tissues in which the protein is expressed. The nucleic
XX CC acid molecules may be used as probes, primers, chemical intermediates and
XX CC in biological assays. These may also be used in diagnosing, preventing or
XX CC treating disorders associated with human transporter protein, such as
XX CC those expressed in foetal heart tissues. The invention is useful in gene
XX CC therapy. The present sequence is human transporter gene.
XX
XX Sequence 34337 BP; 10987 A; 5399 C; 6061 G; 11588 T; 0 U; 302 Other;
XX SQ
XX
XX Query Match 42.5%; Score 27.2; DB 10; Length 34337;
XX Best Local Similarity 72.9%; Pred. No. 42;
XX Matches 35; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
XX
XX 2 ATCCGGAGAGAGAGTTCAGAACGCTGAGAGAGCTGTTAGCAAAAT 49
XX 21458 ATCCCGGAGAGAGCTTAAATATGCCGAGAGAGCTCTGGCATAT 21505
XX
XX RESULT 11
XX ADL00404
XX ID ADL00404 standard; DNA; 78 BP.
XX
XX ADL00404;
XX
XX 06-MAY-2004 (first entry)
XX
XX Antibacterial DNA #1.
XX
XX Antibacterial peptide; plant pathogen resistance; biochemical activity;
XX KM bacterial infection; microbial infection; ds.
XX

OS Unidentified.
XX
XX CN1154970-A.
XX 23-JUL-1997.
XX 16-JAN-1996; 96CN-00100376.
XX 16-JAN-1996; 96CN-00100376.
XX 16-JAN-1996; 96CN-00100376.
XX
XX (BIOL-) BIOLOGICAL TECH RES CENT CHINA AGRIC SCI.
XX Jia S, Qu X, Feng L;
XX WPI: 2004-012928/02.
XX DR
XX New antibiotic peptide used for the control of plant, animal or human
XX PT pathogens and resisting microbial infection.
XX
XX Example 3; SEQ ID NO 5; 63pp; Chinese.
XX
XX The invention relates to a group of new antibacterial peptides that
XX CC possess stronger resistance to plant pathogens than that of natural
XX CC antibacterial peptides, in particular, stronger biochemical activity for
XX CC resisting bacterial infection. The invention also provides the chemical
XX CC synthesizing process and DNA recombination technology for production of
XX CC the antibacterial peptides and the application of the antibacterial
XX CC peptides and their fusion bodies in the control of plant, animal or human
XX CC pathogens and in resisting microbial infection. This sequence represents
XX CC antibacterial DNA of the invention.
XX
XX Sequence 78 BP; 26 A; 14 C; 25 G; 13 T; 0 U; 0 Other;
XX SQ
XX
XX Query Match 41.9%; Score 26.8; DB 12; Length 78;
XX Best Local Similarity 68.5%; Pred. No. 12;
XX Matches 37; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
XX
XX 11 GAAGGTGTTCAAAACCCCTGAGAGAGCTGTTAGCAAAATCTGAGAGAGTA 64
XX 18 GAAGGTGTTCAAAAGATGAGAGAGATGAGAGAGATGAGAGATGAGAGAGCTAGTA 71
XX
XX RESULT 12
XX ADS51259
XX ID ADS51259 standard; cDNA; 1137 BP.
XX
XX ADS51259;
XX
XX 02-DEC-2004 (first entry)
XX
XX Bacterial polynucleotide #6002.
XX
XX Recombinant DNA construct; transformed plant; improved plant property;
XX KM cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
XX KM pathogen tolerance; pest tolerance; plant disease resistance;
XX KM cell cycle pathway modification; plant growth regulator;
XX KM homologous recombination; seed oil yield; protein yield; carbohydrate;
XX KM nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
XX KM bacterial polynucleotide; gene; ss.
XX
XX Bacteria.
XX
XX US2003233675-A1.
XX 18-DEC-2003.
XX 20-FEB-2003; 2003US-00369493.
XX 21-FEB-2002; 2002US-0360039P.
XX
XX (CAOY/) CAO Y.
XX (HINK/) HINKLE G J.
XX (SLAT/) SLATER S C.
XX

PA (CHEN/) CHEN X.
PA (GOLD/) GOLDMAN B S.
XX
XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX WPI; 2004-061375/06.
XX
XX
XX New recombinant DNA construct comprising a promoter positioned to provide
XX PT for expression of a polynucleotide encoding a polypeptide from a
XX PT microbial source, useful for producing plants with improved properties.
XX
XX
XX Claim 1; SEQ ID NO 29689; 122bp; English.
XX
XX The invention relates to a recombinant DNA construct comprising a
XX CC promoter functional in a plant cell, where the promoter is positioned
XX CC to provide for expression of a polynucleotide encoding a polypeptide from a
XX CC microbial source. The invention also relates to a transformed plant
XX CC comprising the recombinant DNA construct and a method of producing a
XX CC transgenic plant having an improved property. The plant is a crop plant
XX CC such as maize or soybean. The method of producing a transformed plant
XX CC having an improved property comprises transforming a plant with the
XX CC recombinant DNA construct and growing the transformed plant, where the
XX CC polynucleotide or polypeptide is useful for improving plant properties.
XX CC The recombinant DNA construct is useful for producing plants with
XX CC improved plant properties, e.g. improved cold, heat or drought tolerance,
XX CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
XX CC increased resistance to plant disease, better growth rate by modification
XX CC of the cell cycle pathway with plant growth regulators, increased rate of
XX CC homologous recombination, modified seed oil or protein yield and/or
XX CC content, improved yield by modification of carbohydrate, nitrogen or
XX CC phosphorus use and/or uptake, by modification of photosynthesis or by
XX CC providing improved plant growth and development under at least one stress
XX CC condition, improved lignin production or improved galactomannan
XX CC production. This sequence represents a bacterial polynucleotide used in
XX CC the scope of the invention. Note: The sequence data for this patent did
XX CC not form part of the printed specification but was obtained in electronic
XX CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 1137 BP; 282 A; 266 C; 278 G; 311 T; 0 U; 0 Other;
XX
Query Match 41.2%; Score 26.4; DB 13; Length 1137;
Best Local Similarity 65.0%; Pred. No. 33;
Matches 39; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
XX
QY 1 GATCCGCAAGAGCTGTTCACACCGCTCGAGAGAGCTGTTGCAAAATCTGGAATCGA 60
DB 429 GATCCAGGAATCAGCTGTTCACACCGCTCTGCAAGCCGAACTCAAGTCAGATCAGCA 488
XX
RESULT 13
ADM03140
ID ADM03140 standard; cDNA; 1578 BP.
XX
XX ADM03140;
XX
XX 20-MAY-2004 (first entry)
XX
XX Human cDNA of the invention SEQ ID NO:1825.
XX
XX ss; gene; human; gene therapy; diagnostic marker; pharmaceutical.
XX
XX Homo sapiens.
XX
XX EP1347046-A1.
XX
XX 24-SEP-2003.
XX
XX 12-APR-2002; 2002EP-00008400.
XX
XX 22-MAR-2002; 2002JP-00137785.
XX
XX (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
XX

PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masubo Y;
XX
XX WPI; 2003-723556/69.
XX
XX P-PSDB; ADM05583.
XX
XX
XX New polynucleotides and polypeptides are useful in gene therapy, for
XX PT developing a diagnostic marker or medicines for regulating their
XX PT expression and activity, or as a target of gene therapy.
XX
XX
XX Claim 1; SEQ ID NO 1825; 305bp; English.
XX
XX
XX The invention relates to a novel human polynucleotide and the encoded
XX CC polypeptide. A polynucleotide of the invention may have a use in gene
XX CC therapy. An oligonucleotide of the invention ADM06202-ADM06773 is useful
XX CC as a primer for synthesizing the polynucleotide or as a probe for
XX CC detecting the polynucleotide. The polynucleotides ADM01316-ADM03758 are
XX CC useful in gene therapy, for developing a diagnostic marker or medicines
XX CC for regulating their expression and activity, or as a target of gene
XX CC therapy. The proteins ADM03759-ADM06201 encoded by the polynucleotides
XX CC are useful as pharmaceutical agents. The present sequence represents a
XX CC cDNA sequence of the invention.
XX
SQ Sequence 1578 BP; 505 A; 320 C; 329 G; 424 T; 0 U; 0 Other;
XX
Query Match 41.2%; Score 26.4; DB 11; Length 1578;
Best Local Similarity 69.2%; Pred. No. 36;
Matches 36; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
XX
QY 11 GAGGTGTTCAACGCGCTGAGAGAGCTGTTCGAAATCTGGAATCGAAG 62
DB 546 GACGATGTTCACGACCTGCGAGGAGGACATTTGAGATGATTCCTGAACCTGAAG 597
XX
RESULT 14
AAT68323/C
ID AAT68323 standard; cDNA; 2681 BP.
XX
XX AAT68323;
XX
XX 09-AUG-1997 (first entry)
XX
XX Clarkia breweri S-linalool synthase cDNA.
XX
XX DE Clarkia breweri S-linalool synthase cDNA.
XX
XX KM S-linalool synthase; scent; fragrance; aroma; flavour; transgenic plant;
XX KM ss.
XX
XX Clarkia breweri.
XX
XX OS
XX FH
XX Key Location/Qualifiers
XX FT CDS 28..2640
XX FT /*tag= a
XX
XX WO9715584-A2.
XX
XX 01-MAY-1997.
XX
XX PD
XX 15-OCT-1996; 96WO-US016482.
XX
XX PF 12-OCT-1995; 95US-0005146P.
XX
XX PR (UNMI) UNIV MICHIGAN.
XX
XX PA
XX Pichersky E;
XX PI
XX WPI; 1997-258947/23.
XX
XX P-PSDB; AAM17080.
XX
XX Linalool synthase from Clarkia breweri (Onograceae) - an acyclic
XX PT mono-terpene, useful for enhancing the scent production or flavour of
XX PT plants.
XX
XX

PS Claim 3; Page 73-82; 105pp; English.

CC A CDNA clone (AA1768323) codes for Clarkia breweri S-linalool synthase
(AA17680), an enzyme that catalyzes the conversion of geranyl.
CC pyrophosphate to S-linalool, an acyclic monoterpene that is a major
CC component of the plant's scent. It was obtained by PCR amplification of cDNA
CC from petal and stigmata cDNA libraries. The clone can be used to produce
CC S-linalool synthase in host cells, to enhance scent production in a
CC transgenic plant such as petunia, rose, carnation etc., or to enhance the
CC flavour of a transgenic plant such as tomato, grape and tea

SO Sequence 2681 BP; 820 A; 607 C; 532 G; 722 T; 0 U; 0 Other;

Query Match 40.6%; Score 26; DB 2; Length 2681;

Best Local Similarity 65.5%; Pred. No. 58;

Matches 38; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 6 GCGAAGAGGTGTTCAACGCTGAGAGAGCTTTAGCAAAATCTGAACTGGAAGT 63
DB 939 GAGTAGAAGGTGTCATCTCCGGATGAGATGTTCACTCAACCCGACCTTGTCAGT 882

RESULT 15

AAV08878/C

ID AAV08878 standard; DNA; 2681 BP.

AC AAV08878;

DT 19-FEB-1999 (first entry)

DE Linalool synthase coding sequence.

KW Linalool synthase; scent enhancer; ss.

OS Clarkia breweri.

Key Location/Qualifiers

FT CDS 28..2640

FT /*tag= a

PN US5849526-A.

PD 15-DEC-1998.

PF 15-OCT-1996; 96US-00732192.

PR 15-OCT-1996; 96US-00732192.

PA (UNMI) UNIV MICHIGAN.

PI Pichersky E;

DR WPI; 1999-069727/06.

DR P-PSDB; AAW73485.

PT DNA encoding Clarkia linalool synthase protein - for producing recombinant protein or transgenic plants.

PS Claim 7; Col 33-40; 30pp; English.

CC This sequence represents the nucleic acid of the invention, encoding the
CC linalool synthase protein of Clarkia breweri. The DNA is used to produce
CC recombinant Clarkia breweri S-linalool synthase or to produce transgenic
CC plants that express Clarkia breweri S-linalool synthase (especially for
CC the production of enhanced scent and taste in plants)

SO Sequence 2681 BP; 820 A; 607 C; 532 G; 722 T; 0 U; 0 Other;

Query Match 40.6%; Score 26; DB 2; Length 2681;

Best Local Similarity 65.5%; Pred. No. 58;

Matches 38; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY 6 GCGAAGAGGTGTTCAACGCTGAGAGAGCTTTAGCAAAATCTGAACTGGAAGT 63

DB

939 GAGTAGAAGGTGTCATCTCCGGATGAGATGTTCACTCAACCCGACCTTGTCAGT 882

Search completed: February 27, 2006, 07:05:28

Job time: 44.6036 secs

THIS PAGE BLANK (0151210)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceeleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:34:28 ; Search time 70.0816 Seconds
(without alignments)
7551.777 Million cell updates/sec

Title: US-10-789-164-6

Perfect score: 64
1 gatccgcgaagaaggtgttc.....aaatctggaactggaagta 64

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Published Applications NA Main:

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	64	100.0	64	9 US-10-789-164-6	Sequence 6, Appl1
2	60	93.8	64	9 US-10-789-164-7	Sequence 7, Appl1
3	28.4	44.4	1659	7 US-10-437-963-86954	Sequence 86954, A
4	27.6	43.1	1706	7 US-10-264-237-1204	Sequence 1204, Ap
5	27.4	42.8	81826	5 US-10-175-523-197	Sequence 197, App
6	27.4	42.8	81826	10 US-11-099-266-197	Sequence 197, App
7	27.2	42.5	554	4 US-09-925-065A-227036	Sequence 227036,
8	27.2	42.5	556	4 US-09-925-065A-171598	Sequence 171598,
9	27.2	42.5	34337	3 US-09-741-149-3	Sequence 3, Appl1
10	27.2	42.5	34337	6 US-10-385-614-3	Sequence 29689, A
11	26.4	41.2	1137	6 US-10-369-493-29689	Sequence 1825, Ap
12	26.4	41.2	1578	6 US-10-108-260A-1825	Sequence 78386, A
13	26.4	41.2	2622	7 US-10-437-963-78386	Sequence 150443,
14	26	40.6	118	8 US-10-425-115-150443	Sequence 33231, A
15	26	40.6	201	8 US-10-741-600-33231	Sequence 1, Appl1
16	26	40.6	2760	7 US-10-469-993-1	Sequence 1, Appl1
17	26	40.6	3708	7 US-10-376-931-1	Sequence 3597, Ap
18	26	40.6	4434	9 US-10-501-282-3597	Sequence 3599, Ap
19	26	40.6	4434	9 US-10-501-282-3599	Sequence 17645, A
20	26	40.6	103660	8 US-10-741-600-17645	Sequence 6651, A
21	26	40.6	1754382	9 US-10-501-282-6651	Sequence 20205, A
22	25.6	40.0	1650	6 US-10-029-386-20205	Sequence 74, Appl1
23	25.6	40.0	2125	6 US-10-115-831-74	Sequence 74, Appl1

24	25.6	40.0	10953	3 US-09-764-846-313	Sequence 313, App
25	25.6	40.0	10953	5 US-10-091-483-313	Sequence 313, App
26	25.6	40.0	98865	3 US-09-770-689A-3	Sequence 3, Appl1
27	25.6	40.0	98865	8 US-10-949-419-3	Sequence 3, Appl1
28	25.6	40.0	495269	7 US-10-398-221-8	Sequence 8, Appl1
29	25.6	40.0	3011208	7 US-10-398-221-2058	Sequence 2058, Ap
30	25.4	39.7	269223	7 US-10-672-787-41	Sequence 41, Appl1
31	25.2	39.4	311	3 US-09-864-761-28288	Sequence 28288, A
32	25.2	39.4	443	3 US-09-864-761-11712	Sequence 11712, A
33	25.2	39.4	580	3 US-09-864-761-15859	Sequence 15859, A
34	25.2	39.4	751	7 US-10-424-599-49013	Sequence 49013, A
35	25.2	39.4	2776	6 US-10-247-671-53	Sequence 53, Appl1
36	25.2	39.4	2911	6 US-10-439-741-25	Sequence 25, Appl1
37	25.2	39.4	3113	6 US-10-252-157-151	Sequence 151, App
38	25.2	39.4	3199	5 US-10-044-090-440	Sequence 440, App
39	25.2	39.4	3199	6 US-10-240-965-267	Sequence 267, App
40	25.2	39.4	4537	6 US-10-240-965-266	Sequence 266, App
41	25.2	39.4	4546	5 US-10-044-090-439	Sequence 439, App
42	25.2	39.4	4601	6 US-10-252-157-150	Sequence 150, App
43	25.2	39.4	4713	3 US-09-814-353-19757	Sequence 19757, A
44	25.2	39.4	4750	8 US-10-723-860-6880	Sequence 6880, Ap
45	25.2	39.4	6048	10 US-11-097-143-24461	Sequence 24461, A

ALIGNMENTS

RESULT 1
US-10-789-164-6
; Sequence 6, Application US/10789164
; Publication No. US20050191720A1
; GENERAL INFORMATION:
; APPLICANT: Sung, M.H. et al.
; TITLE OF INVENTION: SURFACE EXPRESSION METHOD OF PEPTIDES P5 AND ANAL3 USING THE GENE
; FILE REFERENCE: P1574
; CURRENT APPLICATION NUMBER: US/10/789,164
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Kopatentlm 1.71
; SEQ ID NO 6
; LENGTH: 64
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: primer
US-10-789-164-6

Query Match 100.0%; Score 64; DB 9; Length 64;
Best Local Similarity 100.0%; Pred. No. 3.7e-13;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	GATCCGCAAGAAGGTGTTCAACGCTCGAGAGCTGTTAGCAAAATCTGGAATCGA	60
DB	1	GATCCGCAAGAAGGTGTTCAACGCTCGAGAGCTGTTAGCAAAATCTGGAATCGA	60

QY 61 AGTA 64
DB 61 AGTA 64

RESULT 2
US-10-789-164-7/c
; Sequence 7, Application US/10789164
; Publication No. US20050191720A1
; GENERAL INFORMATION:
; APPLICANT: Sung, M.H. et al.
; TITLE OF INVENTION: SURFACE EXPRESSION METHOD OF PEPTIDES P5 AND ANAL3 USING THE GENE
; FILE REFERENCE: P1574
; CURRENT APPLICATION NUMBER: US/10/789,164
; NUMBER OF SEQ ID NOS: 9

```
; SOFTWARE: Koparentin 1.71
; SEQ ID NO 7
; LENGTH: 64
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: primer
US-10-789-164-7

Query Match          93.8%; Score 60; DB 9; Length 64;
Best Local Similarity 100.0%; Pred. No. 1e-11;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 CGCGAAGAGGTGTTCAACGCCCTGAGAGAGCTGTTTACCAAAATCTGGAAGTGA 64
DB 64 CGCGAAGAGGTGTTCAACGCCCTGAGAGAGCTGTTTACCAAAATCTGGAAGTGA 5

RESULT 3
US-10-437-963-86954
; Sequence 86954, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Bardazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 86954
; LENGTH: 1659
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_85948C.1
US-10-437-963-86954

Query Match          44.4%; Score 28.4; DB 7; Length 1659;
Best Local Similarity 66.1%; Pred. No. 8;
Matches 41; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

OY 2 ATCCCGAAGAGGTGTTCAACGCCCTGAGAGAGCTGTTTACCAAAATCTGGAAGTGA 61
DB 814 ATCCCGAAGAGGTGTTTCCCAAGTCGGAGAGAGCATGAGCAACATCTGCTGGTA 873

OY 62 GT 63
DB 874 CT 875

RESULT 4
US-10-264-237-1204
; Sequence 1204, Application US/10264237
; Publication No. US20040009491A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA31PI
; CURRENT APPLICATION NUMBER: US/10/264,237
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/16450
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/205,515
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 2676
```

```
; SOFTWARE: Patentin Ver. 3.1
; SEQ ID NO 1204
; LENGTH: 1706
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-264-237-1204

Query Match          43.1%; Score 27.6; DB 6; Length 1706;
Best Local Similarity 67.2%; Pred. No. 16;
Matches 39; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

OY 4 CGCGAAGAGGTGTTCAACGCCCTGAGAGAGCTGTTTACCAAAATCTGGAAGTGA 61
DB 57 CCATGAGAGAGCATGTTCAACCGCATGAGAGAGCTGCTTAACCAACCGTGGCGAGAA 114

RESULT 5
US-10-175-523-197
; Sequence 197, Application US/10175523
; Publication No. US20030096264A1
; GENERAL INFORMATION:
; APPLICANT: Brockman, Jeffrey
; APPLICANT: Evans, David
; APPLICANT: Hook, Derek
; APPLICANT: Klimczek, Leszek
; APPLICANT: Laeng, Pascal
; APPLICANT: Paley, Michael
; APPLICANT: Rajan, Pritih
; TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHTS)
; FILE REFERENCE: 3235/13795-US3
; CURRENT APPLICATION NUMBER: US/10/175,523
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 60/299,151
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: US 60/317,828
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US 60/325,150
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/333,047
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: US 60/349,936
; PRIOR FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US 60/361,834
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 197
; LENGTH: 81826
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(81826)
; OTHER INFORMATION: where n may be a or g or c or t/u, unknown, or other
US-10-175-523-197

Query Match          42.8%; Score 27.4; DB 5; Length 81826;
Best Local Similarity 69.8%; Pred. No. 65;
Matches 37; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

OY 3 TCCGGAAGAGGTGTTCAACGCCCTGAGAGAGCTGTTTACCAAAATCTGGA 55
DB 6519 TCCGGAAGAGGTGTTTCCGCGCATGAGATCTGATTGAGAAATAGCA 6571

RESULT 6
US-11-099-266-197
; Sequence 197, Application US/11099266
; Publication No. US20050181433A1
; GENERAL INFORMATION:
; APPLICANT: Brockman, Jeffrey
; APPLICANT: Evans, David
; APPLICANT: Hook, Derek
```

APPLICANT: Klineczak, Leszek
APPLICANT: Laeng, Pascal
APPLICANT: Palitayman, Michael
APPLICANT: Rajan, Prithi
TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPPTS)
FILE REFERENCE: 03235/1004795-USA
CURRENT APPLICATION NUMBER: US/11/099,266
PRIOR FILING DATE: 2005-04-04
PRIOR APPLICATION NUMBER: US 10/175,523
PRIOR FILING DATE: 2002-06-18
PRIOR APPLICATION NUMBER: US 60/299,151
PRIOR FILING DATE: 2001-06-18
PRIOR APPLICATION NUMBER: US 60/317,828
PRIOR FILING DATE: 2001-09-07
PRIOR APPLICATION NUMBER: US 60/325,150
PRIOR FILING DATE: 2001-09-25
PRIOR APPLICATION NUMBER: US 60/333,047
PRIOR FILING DATE: 2001-11-14
PRIOR APPLICATION NUMBER: US 60/349,936
PRIOR FILING DATE: 2002-01-18
PRIOR APPLICATION NUMBER: US 60/361,834
PRIOR FILING DATE: 2002-03-04
NUMBER OF SEQ ID NOS: 197
SOFTWARE: PatentIn version 3.1
SEQ ID NO 197
LENGTH: 81826
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(81826)
OTHER INFORMATION: where n may be a or g or c or t/u, unknown, or other
US-11-099-266-197

Query Match 42.8%; Score 27.4; DB 10; Length 81826;
Best Local Similarity 69.8%; Pred. No. 65;
Matches 37; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 3 TCCGCGAAGAGGTGTTCAACGCTGAGAGAGCTGTTAGCAAAATCTGGAA 55
DB 6519 TCCGCGTAGAAGAGATTGTGCGCATGTGATCTGATTGAAATAATGCA 6571

RESULT 7

US-09-925-065A-227036/c
Sequence 227036, Application US/09925065A
Publication No. US20050228172A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single
TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
FILE REFERENCE: 108827,135
CURRENT APPLICATION NUMBER: US/09/925,065A
PRIOR FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: US 60/243,096
PRIOR FILING DATE: 2000-10-24
PRIOR APPLICATION NUMBER: US 60/252,147
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/250,092
PRIOR FILING DATE: 2000-11-30
PRIOR APPLICATION NUMBER: US 60/261,766
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/289,846
PRIOR FILING DATE: 2001-05-09
NUMBER OF SEQ ID NOS: 95086
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 227036
LENGTH: 554
TYPE: DNA
ORGANISM: Homo sapiens
US-09-925-065A-227036

Query Match 42.5%; Score 27.2; DB 4; Length 554;

Best Local Similarity 64.1%; Pred. No. 15;
Matches 41; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 1 GATCCGCAAGAGAGGTGTTCAACGCTGAGAGAGCTGTTAGCAAAATCTGGA 60
DB 236 GCTCCCAAGAGAGATTGTGCGCATGTGATCTGATTGAAATAATGCA 237

QY 61 AGTA 64
DB 236 GGTA 233

RESULT 8

US-09-925-065A-171598/c
Sequence 171598, Application US/09925065A
Publication No. US20050228172A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single
TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
FILE REFERENCE: 108827,135
CURRENT APPLICATION NUMBER: US/09/925,065A
PRIOR FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: US 60/243,096
PRIOR FILING DATE: 2000-10-24
PRIOR APPLICATION NUMBER: US 60/252,147
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/250,092
PRIOR FILING DATE: 2000-11-30
PRIOR APPLICATION NUMBER: US 60/261,766
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/289,846
PRIOR FILING DATE: 2001-05-09
NUMBER OF SEQ ID NOS: 957086
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 171598
LENGTH: 556
TYPE: DNA
ORGANISM: Homo sapiens
US-09-925-065A-171598

Query Match 42.5%; Score 27.2; DB 4; Length 556;
Best Local Similarity 64.1%; Pred. No. 15;
Matches 41; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 1 GATCCGCAAGAGAGGTGTTCAACGCTGAGAGAGCTGTTAGCAAAATCTGGA 60
DB 236 GCTCCCAAGAGAGATTGTGCGCATGTGATCTGATTGAAATAATGCA 237

QY 61 AGTA 64
DB 236 GGTA 233

RESULT 9

US-09-741-149-3
Sequence 3, Application US/09741149
Patent No. US20020031800A1
GENERAL INFORMATION:
APPLICANT: Li, Zhenya et al
TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
FILE REFERENCE: CLO00780
CURRENT APPLICATION NUMBER: US/09/741,149
PRIOR FILING DATE: 2000-12-21
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 34337
TYPE: DNA
ORGANISM: Human
US-09-741-149-3

Query Match 42.5%; Score 27.2; DB 4; Length 556;

NAME/KEY: misc.feature
LOCATION: (1)_(34337)
OTHER INFORMATION: n = A,T,C or G
US-09-741-149-3

Query Match 42.5%; Score 27.2; DB 3; Length 34337;
Best Local Similarity 72.9%; Pred. No. 58;
Matches 35; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 2 ATCCGCGAAGAGGTGTTCAACGCGCTGGAGAGCTGTTAGCAAAAT 49
Db 21458 ATGCCCGAAGAGCTTTAAATATGCCGAGAGCTCCTGGCATTAT 21505

RESULT 10
US-10-385-614-3
; Sequence 3, Application US/10385614
; Publication No. US20030157649A1
; GENERAL INFORMATION:
; APPLICANT: Li, Zhenya et al
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; FILE REFERENCE: CLO00780
; CURRENT APPLICATION NUMBER: US/10/385,614
; CURRENT FILING DATE: 2003-03-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FASTSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 34337
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(34337)
; OTHER INFORMATION: n = A,T,C or G
US-10-385-614-3

Query Match 42.5%; Score 27.2; DB 6; Length 34337;
Best Local Similarity 72.9%; Pred. No. 58;
Matches 35; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 2 ATCCGCGAAGAGGTGTTCAACGCGCTGGAGAGCTGTTAGCAAAAT 49
Db 21458 ATGCCCGAAGAGCTTTAAATATGCCGAGAGCTCCTGGCATTAT 21505

RESULT 11
US-10-369-493-29689
; Sequence 29689, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 29689
; LENGTH: 1137
; TYPE: DNA
; ORGANISM: Caenorhabditis elegans
US-10-369-493-29689

Query Match 41.2%; Score 26.4; DB 6; Length 1137;
Best Local Similarity 65.0%; Pred. No. 37;

Matches 39; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Qy 1 GATCCGCGAAGAGGTGTTCAACGCGCTGGAGAGCTGTTAGCAAAATCGAAGTGA 60
Db 429 GATCCGAGATCAGCTGTTCACAGCGCTTGGACAGCGGACATCAAGCTCAGATCAGA 488

RESULT 12
US-10-108-260A-1825
; Sequence 1825, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20040005560A1el full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1825
; LENGTH: 1578
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-108-260A-1825

Query Match 41.2%; Score 26.4; DB 6; Length 1578;
Best Local Similarity 69.2%; Pred. No. 41;
Matches 36; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 11 GAAGGTGTTCAACGCGCTGGAGAGCTGTTAGCAAAATCGAAGTGAAG 62
Db 546 GACGATGTTCAAGCAGCTGGAGAGCAATTGAGATGATTCCTGAATTTGAAG 597

RESULT 13
US-10-437-963-78386
; Sequence 78386, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Mu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 78386
; LENGTH: 2622
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_78199C.1
US-10-437-963-78386

Query Match 41.2%; Score 26.4; DB 7; Length 2622;
Best Local Similarity 65.0%; Pred. No. 49;
Matches 39; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Qy 5 CGCGAAGAGGTGTTCAACGCGCTGGAGAGCTGTTAGCAAAATCGAAGTGAAGTA 64
Db 359 CGTGTGAAGCTGTTCAACGAGCTGGAGAGCCGCTGCTTCACTGGAAGGAGATCA 418

RESULT 14
US-10-425-115-150443
; Sequence 150443, Application US/10425115

```
Publication No. US20040214272A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21 (5322)B
CURRENT APPLICATION NUMBER: US/10/425,115
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 369326
SEQ ID NO 150443
LENGTH: 118
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: MRT4577_6872C.1
US-10-425-115-150443

Query Match          40.6%; Score 26; DB 8; Length 118;
Best Local Similarity 70.0%; Pred. No. 25;
Matches 35; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 6 GCCAAGAGGTGTTCAAGCGCTGAGAGAGCTTTAGCAAAATCTGGAA 55
DB 45 GCCAGAGAGGTGTTCAAGCGCTGAGAGAGCTTTAGCAAAATCTGGAA 94

RESULT 15
US-10-741-600-33231/C
Sequence 33231, Application US/10741600
Publication No. US20050026169A1
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele et al.
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
FILE REFERENCE: CL001499
CURRENT APPLICATION NUMBER: US/10/741,600
CURRENT FILING DATE: 2003-12-22
NUMBER OF SEQ ID NOS: 73997
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 33231
LENGTH: 201
TYPE: DNA
ORGANISM: Homo sapiens
US-10-741-600-33231

Query Match          40.6%; Score 26; DB 8; Length 201;
Best Local Similarity 70.0%; Pred. No. 30;
Matches 35; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 8 GAGGAAGCTGTTCAAGCGCTGAGAGAGCTTTAGCAAAATCTGGAACT 57
DB 73 GATGAAGATGTTCAAGCGCTGAGAGAGCTTTAGCAAAATCTGGAACT 24
```

Search completed: February 27, 2006, 07:58:48
Job time : 74.0816 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceeleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:37:00 ; Search time 38.5681 Seconds
(without alignments)
3538.457 Million cell updates/sec

Title: US-10-789-164-6

Perfect score: 64
Sequence: 1 gatccgcgaagaagtgcttc.....aaatctggaactggaagta 64

Scoring table: IDENTITY NUC
Gapop 10_0, Gapext 1.0

Searched: 7209121 seqs, 1066183437 residues

Total number of hits satisfying chosen parameters: 14418242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Published Applications NA New.*

1: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
2: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	27.2	42.5	554	US-09-925-065A-227036	Sequence 227036, A
C 2	27.2	42.5	556	US-09-925-065A-171598	Sequence 171598, A
C 3	26	40.6	201	US-10-995-561-31012	Sequence 31012, A
C 4	26	40.6	103660	US-10-995-561-13253	Sequence 13253, A
C 5	24.8	38.8	643	US-09-925-065A-897354	Sequence 897354, A
C 6	24.8	38.8	10302	US-11-127-832-23	Sequence 23, Appl
C 7	24.4	38.1	201	US-10-995-561-64267	Sequence 64267, A
C 8	24.4	38.1	201	US-10-995-561-64483	Sequence 64483, A
C 9	24.4	38.1	523	US-09-925-065A-312415	Sequence 312415, A
C 10	24.4	38.1	852	US-09-925-065A-59275	Sequence 59275, A
C 11	24.2	37.8	201	US-11-124-368A-11876	Sequence 11876, A
C 12	24.2	37.8	2416	US-09-925-065A-709999	Sequence 709999, A
C 13	24.2	37.8	2416	US-09-925-065A-710000	Sequence 710000, A
C 14	24.2	37.8	2416	US-09-925-065A-710001	Sequence 710001, A
C 15	24	37.5	201	US-10-995-561-64268	Sequence 64268, A
C 16	24	37.5	382	US-11-031-356-24	Sequence 24, Appl
C 17	24	37.5	519	US-09-925-065A-225239	Sequence 225239, A
C 18	24	37.5	584	US-09-925-065A-608992	Sequence 608992, A
C 19	24	37.5	2437	US-11-072-512-1159	Sequence 1159, Appl
C 20	24	37.5	56448	US-10-995-561-13369	Sequence 13369, A

21	23.8	37.2	201	US-11-124-368A-11820	Sequence 11820, A
22	23.8	37.2	569	US-09-925-065A-269497	Sequence 269497, A
23	23.8	37.2	569	US-09-925-065A-269498	Sequence 269498, A
C 24	23.8	37.2	1922	US-10-750-185-38496	Sequence 38496, A
C 25	23.8	37.2	1922	US-10-750-623-38496	Sequence 38496, A
C 26	23.6	36.9	741	US-09-925-065A-66253	Sequence 66253, A
C 27	23.6	36.9	1159	US-10-750-185-27344	Sequence 27344, A
C 28	23.6	36.9	1159	US-10-750-623-37344	Sequence 37344, A
C 29	23.6	36.9	1468	US-09-925-065A-698874	Sequence 698874, A
C 30	23.6	36.9	2064	US-11-117-169-7	Sequence 7, Appl
C 31	23.6	36.9	2186	US-11-117-169-14	Sequence 1, Appl
C 32	23.6	36.9	3068	US-11-117-169-14	Sequence 75, Appl
C 33	23.6	36.9	5032	US-10-775-169-75	Sequence 37, Appl
C 34	23.6	36.9	5032	US-10-947-249-37	Sequence 586975, A
C 35	23.4	36.6	442	US-09-925-065A-586975	Sequence 671345, A
C 36	23.4	36.6	656	US-09-925-065A-671345	Sequence 308202, A
C 37	23.2	36.2	596	US-09-925-065A-308202	Sequence 681453, A
C 38	23.2	36.2	1318	US-09-925-065A-681453	Sequence 681454, A
C 39	23.2	36.2	1318	US-09-925-065A-681454	Sequence 21, Appl
C 40	23.2	36.2	175100	US-11-121-086-21	Sequence 11889, A
C 41	23	35.9	201	US-11-124-368A-11889	Sequence 77828, A
C 42	23	35.9	1167	US-09-925-065A-77828	Sequence 12, Appl
C 43	23	35.9	1773	US-11-040-218-12	Sequence 1670, Appl
C 44	23	35.9	1847	US-11-072-512-1670	Sequence 77, Appl
C 45	23	35.9	2557	US-10-775-169-77	

ALIGNMENTS

RESULT 1
US-09-925-065A-227036/C
Sequence 227036, Application US/09925065A
Publication No. US2004081048A1
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single
TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
FILE REFERENCE: 108827.135
CURRENT APPLICATION NUMBER: US/09/925, 065A
CURRENT FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: US 60/243, 096
PRIOR FILING DATE: 2000-10-24
PRIOR APPLICATION NUMBER: US 60/252, 147
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/250, 092
PRIOR FILING DATE: 2000-11-30
PRIOR APPLICATION NUMBER: US 60/261, 766
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/289, 846
NUMBER OF SEQ ID NOS: 957086
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 227036
LENGTH: 554
TYPE: DNA
ORGANISM: Homo sapiens
US-09-925-065A-227036

Query Match 42.5%; Score 27.2; DB 6; Length 554;
Best Local Similarity 64.1%; Pred. No. 3.1;
Matches 41; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 1 GATCCGCAAGAAGGTTTCAACGCTGAGAGAGCTGTTTACCAAAATCTGGAATCGA 60
DB GCTCCCAAGAAGATTGTCAGGCTCATGAGATCTTGAACCAAAATTCGAATTGGA 237

QY 61 AGTA 64

DB 236 GGA 233

RESULT 2

```
US-09-925-065A-171598/c
; Sequence 171598, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 171598
; LENGTH: 556
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-171598
```

```
Query Match 42.5%; Score 27.2; DB 6; Length 556;
Best Local Similarity 64.1%; Pred. No. 3.1;
Matches 41; Conservative 0; Mismatches 23; Indels 0; Gaps 0;
```

```
Oy 1 GATCGGCAAGAGGTGTTCAACGCCCTGGAGAACTGTTAGCAAAATCTGGAACCTGA 60
Db 236 GCTCCACAGAGATTGGTTCAGGCTCATGAGAACTTTGAACCAAAATTCGAATTGGA 237
Oy 61 AGTA 64
Db 236 GGTA 233
```

```
RESULT 3
US-10-995-561-31012/c
; Sequence 31012, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: C1001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31012
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-31012
```

```
Query Match 40.6%; Score 26; DB 8; Length 201;
Best Local Similarity 70.0%; Pred. No. 6.4;
Matches 35; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
```

```
Oy 8 GAAGAAGGTGTTCAACGCCCTGGAGAACTGTTTAGCAAAATCTGGAACCTGA 57
Db 73 GATGAAATGTTCAACCATCTTGAAGAGCACTTTGGCAATATCTAGCACT 24
```

```
RESULT 4
US-10-995-561-13253/c
; Sequence 13253, Application US/10995561
```

```
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: C1001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13253
; LENGTH: 103660
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(103660)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-;
US-10-995-561-13253
```

```
Query Match 40.6%; Score 26; DB 8; Length 103660;
Best Local Similarity 70.0%; Pred. No. 31;
Matches 35; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
```

```
Oy 8 GAAGAAGGTGTTCAACGCCCTGGAGAACTGTTTAGCAAAATCTGGAACCTGA 57
Db 47494 GATGAAATGTTCAACCATCTTGAAGAGCACTTTGGCAATATCTAGCACT 47445
```

```
RESULT 5
US-09-925-065A-897354
; Sequence 897354, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 897354
; LENGTH: 643
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-897354
```

```
Query Match 38.8%; Score 24.8; DB 6; Length 643;
Best Local Similarity 67.3%; Pred. No. 23;
Matches 35; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
```

```
Oy 9 AAGAAGGTGTTCAACGCCCTGGAGAACTGTTTAGCAAAATCTGGAACCTGA 60
Db 530 AACTAACTGTCATTAAGCTTGGAGAACTGTGTGAAAGCAAGAACTGAA 581
```

```
RESULT 6
US-11-127-832-23
; Sequence 23, Application US/11127832
; Publication No. US20060008884A1
; GENERAL INFORMATION:
```

```

APPLICANT: Hearing, Patrick
APPLICANT: Bahou, Wadie
APPLICANT: Sandeaton, Ziv
APPLICANT: Gnatenko, Dmitri
TITLE OF INVENTION: Adenoviral Vectors
FILE REFERENCE: STONY-04970
CURRENT APPLICATION NUMBER: US/11/127,832
PRIORITY FILING DATE: 2005-05-12
PRIOR FILING DATE: 2001-02-12
PRIOR APPLICATION NUMBER: 60/237,747
PRIOR FILING DATE: 2000-10-02
NUMBER OF SEQ ID NOS: 27
SOFTWARE: PatentIn version 3.0
SEQ ID NO 23
LENGTH: 10302
TYPE: DNA
ORGANISM: Homo sapiens
US-11-127-832-23
```

Query Match	38.8%;	Score 24.8;	DB 12;	Length 10302;
Best Local Similarity	63.3%;	Pred. No. 47;		
Matches 38;	Conservative	0;	Mismatches 22;	Indels 0;

QY 4 CCGCGAAGAGGCTTCAAACGCTGGAGAGCTGTTAGCAAAATCTGGAATCGGAAGT 63
Db 2652 CCGCTACCAAGCTGTACAAGAGCTGTAGAGATCTGTACAACAATCTAGAGATGAAGT 2711

RESULT 7
US-10-995-561-64267/c
; Sequence 64267, Application US/10995561
; Publication No. US20050272054A1

```

? APPLICANT CARGILL, Michele et al.
? TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
? TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
? TITLE OF INVENTION: DETECTION AND USES THEREOF
? FILE REFERENCE: C1001559
? CURRENT APPLICATION NUMBER: US/10/995,561
? CURRENT FILING DATE: 2004-11-24
? NUMBER OF SEQ ID NOS: 85702
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 64267
? LENGTH: 201
? TYPE: DNA
? ORGANISM: Homo sapiens
? US-10-995-561-64267

```

Query Match	38.1%	Score 24.4;	DB 8;	Length 201;
Best Local Similarity	68.0%;	Pred. No. 24;		
Matches 34;	Conservative 0;	Mismatches 16;	Indels 0;	Gaps 0

```

Oy      9  AAGAGGTCTTCAACCGCTGAGAACTGTTTAGCAAAATCTGGAAC 58
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db     135  AAGAAGATTGCTCAGCATTTGAGTAAATGTTCAACAAATCTGAATTG 866

```

```

RESULT 8
US-10-995-561-64483/c
; Sequence:64483, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 64483

```

```

; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-64483

```

Query Match	38.1%	Score 24.4;	DB 8;	Length 201;
Best Local Similarity	68.0%	Pred. No. 24;		
Matches 34; Conservative	0;	Mismatches 16;	Indels 0;	Gaps 0;

```

Oy      9  AGAAGTGTTCAAACGCCCTGGAGAACTGTTTAGCAAAATCTGGAAGT 58
        |||||
Db      64  AAGAAGATTGCTCAGGCAATTGGTAATGTTTCATCAAAATCTGAATTG 15

```

RESULT 9
US-09-925-065A-312415/c
; Sequence 312415, Application US/09925065A
; Publication No. US20040181048A1
***** INFORMATION *****

```

? TITLE OF INVENTION: Identification and Mapping of Single
? FILE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
? FILE REFERENCE: 108827.135
? CURRENT APPLICATION NUMBER: US/09/925, 065A
? PRIOR FILING DATE: 2001-08-08
? PRIOR APPLICATION NUMBER: US 60/243, 036
? PRIOR FILING DATE: 2000-10-24
? PRIOR APPLICATION NUMBER: US 60/255, 147
? PRIOR FILING DATE: 2000-11-20
? PRIOR APPLICATION NUMBER: US 60/250, 092
? PRIOR FILING DATE: 2000-11-30
? PRIOR APPLICATION NUMBER: US 60/261, 766
? PRIOR FILING DATE: 2001-01-16
? PRIOR APPLICATION NUMBER: US 60/289, 846
? PRIOR FILING DATE: 2001-05-09
? NUMBER OF SEQ ID NOS: 957086
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 312415
? LENGTH: 523
? TYPE: DNA
? ORGANISM: Homo sapiens
US-09-925-065A-312415

```

Query Match	38.1%	Score 24.4	DB 6	length 523
Best Local Similarity	68.0%	Pred. No. 31		
Matches 34	Conservative 0	Mismatches 16	Indels 0	Gaps 0

```
Oy      1 GATCCGGAAGAGGTGTTCAACGCCCTGAGAAGCGTTCAGCAAATC 50
          ||| | | | | | | | | | | | | | | | | | | | |
Db     211 GAGATGGGGGAAGGTGCTCAAAAGCCAAGAGACCATTTCAGCTAAAGC 162
```

```

: RESULT 10
: US-09-925-065A-59275
: Sequence 59275, Application US/09925065A
: Publication No. US20040181048A1
: GENERAL INFORMATION:
: APPLICANT: Wang, David G.
: TITLE OF INVENTION: Identification and Mapping of Single
: TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
: FILE REFERENCE: 108827.135
: CURRENT APPLICATION NUMBER: US/09/925, 065A
: CURRENT FILING DATE: 2001-08-08
: PRIOR APPLICATION NUMBER: US 60/243, 036
: PRIOR FILING DATE: 2000-10-24
: PRIOR APPLICATION NUMBER: US 60/252, 147
: PRIOR FILING DATE: 2000-11-20
: PRIOR APPLICATION NUMBER: US 60/250, 092
: PRIOR FILING DATE: 2000-11-30
: PRIOR APPLICATION NUMBER: US 60/261, 766
: PRIOR FILING DATE: 2001-01-16
: PRIOR APPLICATION NUMBER: US 60/289, 846
: PRIOR FILING DATE: 2001-05-09

```

NUMBER OF SEQ ID NOS: 957086
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 59275
LENGTH: 852
TYPE: DNA
ORGANISM: Homo sapiens
US-09-925-065A-59275

Query Match 38.1%; Score 24.4; DB 6; Length 852;
Best Local Similarity 63.8%; Pred. No. 35;
Matches 37; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Qy 4 CCGGAGAGAGAGTGTTCAAACGCCCTGGAGAGCTGTTAGCAAAATCTGAACTGAA 61
Db 663 CAGTGAATTAAGTGAATTAAGATTGGATATACAGTTAGTGAATATAGAAATGGA 720

RESULT 11
US-11-124-368A-11876
Sequence 11876, Application US/11124368A
Publication No. US20050287559A1
GENERAL INFORMATION:
APPLICANT: Michele Cargili
APPLICANT: James J. Devlin
APPLICANT: May Luke
TITLE OF INVENTION: Genetic Polymorphisms Associated with
FILE REFERENCE: CLO01524
CURRENT APPLICATION NUMBER: US/11/124.368A
CURRENT FILING DATE: 2005-05-09
PRIOR APPLICATION NUMBER: US 60/568,845
PRIOR FILING DATE: 2004-05-07
PRIOR APPLICATION NUMBER: US 60/625,936
PRIOR FILING DATE: 2004-11-09
NUMBER OF SEQ ID NOS: 21112
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11876
LENGTH: 201
TYPE: DNA
ORGANISM: Homo sapiens
US-11-124-368A-11876

Query Match 37.8%; Score 24.2; DB 12; Length 201;
Best Local Similarity 66.0%; Pred. No. 29;
Matches 35; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy 11 GAAGGTGTTCAAACGCCCTGGAGAGCTGTTAGCAAAATCTGAACTGAACT 63
Db 2 GATGAGGTTCAGACTCATGAGAGAAACAGAGGGTGAATTCGATCTGAAAGT 54

RESULT 12
US-09-925-065A-709999/c
Sequence 709999, Application US/09925065A
Publication No. US20040181048A1
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single
FILE REFERENCE: 108827.135
CURRENT APPLICATION NUMBER: US/09/925,065A
CURRENT FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: US 60/243,096
PRIOR FILING DATE: 2000-10-24
PRIOR APPLICATION NUMBER: US 60/252,147
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/250,092
PRIOR FILING DATE: 2000-11-30
PRIOR APPLICATION NUMBER: US 60/261,766
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/289,846
PRIOR FILING DATE: 2001-05-09
NUMBER OF SEQ ID NOS: 957086

SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 709999
LENGTH: 2416
TYPE: DNA
ORGANISM: Homo sapiens
US-09-925-065A-709999

Query Match 37.8%; Score 24.2; DB 6; Length 2416;
Best Local Similarity 71.1%; Pred. No. 53;
Matches 32; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 17 GTTCAACGCCCTGGAGAGCTGTTAGCAAAATCTGAACTGAA 61
Db 878 GTTCAACGCCCTGGAGAGCTGTCATCACTTCAACAGAGAA 834

RESULT 13
US-09-925-065A-710000/c
Sequence 710000, Application US/09925065A
Publication No. US20040181048A1
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single
FILE REFERENCE: 108827.135
CURRENT APPLICATION NUMBER: US/09/925,065A
CURRENT FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: US 60/243,096
PRIOR FILING DATE: 2000-10-24
PRIOR APPLICATION NUMBER: US 60/252,147
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/250,092
PRIOR FILING DATE: 2000-11-30
PRIOR APPLICATION NUMBER: US 60/261,766
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: US 60/289,846
PRIOR FILING DATE: 2001-05-09
NUMBER OF SEQ ID NOS: 957086
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 710000
LENGTH: 2416
TYPE: DNA
ORGANISM: Homo sapiens
US-09-925-065A-710000

Query Match 37.8%; Score 24.2; DB 6; Length 2416;
Best Local Similarity 71.1%; Pred. No. 53;
Matches 32; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 17 GTTCAACGCCCTGGAGAGCTGTTAGCAAAATCTGAACTGAA 61
Db 878 GTTCAACGCCCTGGAGAGCTGTCATCACTTCAACAGAGAA 834

RESULT 14
US-09-925-065A-710001/c
Sequence 710001, Application US/09925065A
Publication No. US20040181048A1
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single
FILE REFERENCE: 108827.135
CURRENT APPLICATION NUMBER: US/09/925,065A
CURRENT FILING DATE: 2001-08-08
PRIOR APPLICATION NUMBER: US 60/243,096
PRIOR FILING DATE: 2000-10-24
PRIOR APPLICATION NUMBER: US 60/252,147
PRIOR FILING DATE: 2000-11-20
PRIOR APPLICATION NUMBER: US 60/250,092
PRIOR FILING DATE: 2000-11-30
PRIOR APPLICATION NUMBER: US 60/261,766
PRIOR FILING DATE: 2001-01-16

PRIOR APPLICATION NUMBER: US 60/289,846
 PRIOR FILING DATE: 2001-05-09
 NUMBER OF SEQ ID NOS: 957086
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 710001
 LENGTH: 2416
 TYPE: DNA
 ORGANISM: Homo sapiens
 US-09-925-065A-710001

Query Match 37.8%; Score 24.2; DB 6; Length 2416;
 Best Local Similarity 71.1%; Pred. No. 53;
 Matches 32; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 17 GTTCAACGCTGAGAACCTGTTAGCAAAATCTGGAATGGA 61
 Db 878 GTTCAACGCTGAGAACCTGTTAGCAAAATCTGGAATGGA 834

RESULT 15
 US-10-995-561-64268/c
 Sequence 64268, Application US/10995561
 Publication No. US20050272054A1
 GENERAL INFORMATION:
 APPLICANT: CARGILL, Michele et al.
 TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
 TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
 TITLE OF INVENTION: DETECTION AND USES THEREOF
 FILE REFERENCE: CL001559
 CURRENT APPLICATION NUMBER: US/10/995,561
 CURRENT FILING DATE: 2004-11-24
 NUMBER OF SEQ ID NOS: 85702
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 64268
 LENGTH: 201
 TYPE: DNA
 ORGANISM: Homo sapiens
 US-10-995-561-64268

Query Match 37.5%; Score 24; DB 8; Length 201;
 Best Local Similarity 66.0%; Pred. No. 34;
 Matches 33; Conservative 1; Mismatches 16; Indels 0; Gaps 0;

Qy 9 AAGAAGTGTTCACACGCTGAGAACCTGTTAGCAAAATCTGGAATG 58
 Db 106 AAGAAGTGTTCACACGCTGAGAACCTGTTAGCAAAATCTGGAATG 57

Search completed: February 27, 2006, 11:38:29
 Job time : 38.5681 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceeleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:34:23 ; Search time 15.0601 Seconds
(without alignments) 7553.984 Million cell updates/sec

Title: US-10-789-164-6

Perfect score: 64
Sequence: 1 gatccgcgaagaaggtgttc.....aaatctggaactgaagta 64

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:*

1: /cgn2_6/prodata/1/ina/1 COMB.seq:*
2: /cgn2_6/prodata/1/ina/5 COMB.seq:*
3: /cgn2_6/prodata/1/ina/6A COMB.seq:*
4: /cgn2_6/prodata/1/ina/6B COMB.seq:*
5: /cgn2_6/prodata/1/ina/H COMB.seq:*
6: /cgn2_6/prodata/1/ina/PTTUS COMB.seq:*
7: /cgn2_6/prodata/1/ina/PP COMB.seq:*
8: /cgn2_6/prodata/1/ina/RP COMB.seq:*
9: /cgn2_6/prodata/1/ina/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
C 1	26	40.6	2681 2	US-08-732-192A-1 Sequence 1, Appl
C 2	26	40.6	2681 3	US-09-172-339-7 Sequence 7, Appl
C 3	25.4	39.7	834 3	US-09-540-236-945 Sequence 945, App
4	25.4	39.7	269223 3	US-09-596-002-41 Sequence 41, App
5	25.2	39.4	601 3	US-09-949-016-25005 Sequence 25005, A
6	25.2	39.4	601 3	US-09-949-016-38480 Sequence 38480, A
7	25.2	39.4	601 3	US-09-949-016-170705 Sequence 170705, A
8	25.2	39.4	601 3	US-09-949-016-170748 Sequence 170748, A
9	25.2	39.4	601 3	US-09-949-016-170791 Sequence 170791, A
10	25.2	39.4	601 3	US-09-949-016-170832 Sequence 170832, A
11	25.2	39.4	601 3	US-09-949-016-176054 Sequence 176054, A
12	25.2	39.4	3416 3	US-09-949-016-4827 Sequence 4827, Ap
13	25.2	39.4	3416 3	US-09-949-016-4828 Sequence 4828, Ap
14	25.2	39.4	3490 3	US-09-949-016-4829 Sequence 4829, Ap
15	25.2	39.4	3490 3	US-09-949-016-4830 Sequence 4830, Ap
16	25.2	39.4	12241 3	US-09-949-016-16746 Sequence 16746, A
17	25.2	39.4	12243 3	US-09-949-016-12094 Sequence 12094, A
18	25.2	39.4	25230 3	US-09-949-016-13153 Sequence 13153, A
19	25.2	39.4	36075 3	US-09-949-016-16571 Sequence 16571, A
20	25.2	39.4	36075 3	US-09-949-016-16572 Sequence 16572, A
21	25.2	39.4	36625 3	US-09-949-016-12788 Sequence 12788, A
22	25.2	39.4	37133 3	US-09-949-016-16569 Sequence 16569, A
23	25.2	39.4	37133 3	US-09-949-016-16570 Sequence 16570, A
24	25	39.1	4174 3	US-09-949-002-138 Sequence 138, App

25	25	39.1	4180 3	US-09-814-915A-93 Sequence 93, Appl
26	25	39.1	4180 3	US-09-949-002-103 Sequence 103, App
27	25	39.1	15044 3	US-09-949-002-675 Sequence 675, App
28	25	39.1	15044 3	US-09-949-002-710 Sequence 710, App
29	24.8	38.8	601 3	US-09-949-016-25006 Sequence 25006, A
30	24.8	38.8	601 3	US-09-949-016-176055 Sequence 176055, A
31	24.8	38.8	10302 3	US-10-149-736-3 Sequence 3, Appl
32	24.8	38.8	10302 3	US-09-782-378A-23 Sequence 23, Appl
33	24.8	38.8	10320 3	US-09-091-501B-9 Sequence 9, Appl
34	24.6	38.4	3514 3	US-09-716-964B-117 Sequence 117, App
35	24.6	38.4	23456 3	US-09-949-016-12989 Sequence 12989, A
36	24.6	38.4	23458 3	US-09-949-016-12605 Sequence 12605, A
37	24.4	38.1	1853 3	US-08-687-590-61 Sequence 61, Appl
38	24.2	37.8	601 3	US-09-949-016-166961 Sequence 166961, A
39	24.2	37.8	601 3	US-09-328-352-1830 Sequence 1830, Ap
40	24.2	37.8	771 3	US-09-949-016-12796 Sequence 12796, A
41	24.2	37.8	30271 3	US-09-949-016-14006 Sequence 14006, A
42	24.2	37.8	30272 3	US-09-949-016-16436 Sequence 16436, A
43	24.2	37.8	33125 3	US-09-949-016-16437 Sequence 16437, A
44	24.2	37.8	33125 3	US-09-949-016-16437 Sequence 16437, A
45	24	37.5	382 3	US-09-284-782-24 Sequence 24, Appl

ALIGNMENTS

RESULT 1
US-08-732-192A-1/c
; Sequence 1, Application US/08732192A
; Patent No. 5849526
; GENERAL INFORMATION:
; APPLICANT: Pichersky, Brian
; TITLE OF INVENTION: USE OF LINALOOL SYNTHASE IN GENETIC
; TITLE OF INVENTION: ENGINEERING OF SCENT PRODUCTION
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/732,192A
; FILING DATE: 15-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/005,146
; FILING DATE: 25-OCT-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: HIGHLANDER, STEVEN L.
; REGISTRATION NUMBER: 37,642
; REFERENCE/DOCKET NUMBER: UMIC:015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (713) 789-2679
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2681 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 28..2637
; US-08-732-192A-1
Query Match * 40.6%; Score 26; DB 2; Length 2681;

RESULT 6
US-09-949-016-38480
; Sequence 38480, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 38480
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-38480

Query Match 39.4%; Score 25.2; DB 3; Length 601;
Best Local Similarity 71.7%; Pred. No. 15;
Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 GATCCGGAAGAGCTTCAAAACGCTGAGAGAGCTGTTAGCA 46
DB 294 GAACCGTCTGAGCTGATCAATGCGAGAGAGCTGTATCTA 339

RESULT 7
US-09-949-016-170705
; Sequence 170705, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 170705
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-170705

Query Match 39.4%; Score 25.2; DB 3; Length 601;
Best Local Similarity 71.7%; Pred. No. 15;
Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 GATCCGGAAGAGCTTCAAAACGCTGAGAGAGCTGTTAGCA 46
DB 294 GAACCGTCTGAGCTGATCAATGCGAGAGAGCTGTATCTA 339

RESULT 8
US-09-949-016-170748
; Sequence 170748, Application US/09949016
; Patent No. 6812339

; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 170748
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-170748

Query Match 39.4%; Score 25.2; DB 3; Length 601;
Best Local Similarity 71.7%; Pred. No. 15;
Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 GATCCGGAAGAGCTTCAAAACGCTGAGAGAGCTGTTAGCA 46
DB 294 GAACCGTCTGAGCTGATCAATGCGAGAGAGCTGTATCTA 339

RESULT 9
US-09-949-016-170791
; Sequence 170791, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 170791
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-170791

Query Match 39.4%; Score 25.2; DB 3; Length 601;
Best Local Similarity 71.7%; Pred. No. 15;
Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 GATCCGGAAGAGCTTCAAAACGCTGAGAGAGCTGTTAGCA 46
DB 294 GAACCGTCTGAGCTGATCAATGCGAGAGAGCTGTATCTA 339

RESULT 10
US-09-949-016-170832
; Sequence 170832, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307

```
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 170832
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-170832

Query Match          39.4%; Score 25.2; DB 3; Length 601;
Best Local Similarity 71.7%; Pred. No. 15;
Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 GATCCGGAAGAGGTGTTCAAGCGCTGGAGAGCTGTTAGCAA 46
DB 294 GAACCGTCTGAGCTGATCAATGCCAGAGAGCTGTATCTA 339

RESULT 11
US-09-949-016-176054
; Sequence 176054, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 176054
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-176054

Query Match          39.4%; Score 25.2; DB 3; Length 601;
Best Local Similarity 71.7%; Pred. No. 15;
Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 3 TCCGCGAAGAGGTGTTCAAGCGCTGGAGAGCTGTTAGCAA 48
DB 225 TCCGGGTAGAGGATTTGTGCGCATGTGATCTGATTAGAAAA 280

RESULT 12
US-09-949-016-4827
; Sequence 4827, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
```

```
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4827
; LENGTH: 3416
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-4827

Query Match          39.4%; Score 25.2; DB 3; Length 3416;
Best Local Similarity 71.7%; Pred. No. 26;
Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 GATCCGGAAGAGGTGTTCAAGCGCTGGAGAGCTGTTAGCAA 46
DB 765 GAACCGTCTGAGCTGATCAATGCCAGAGAGCTGTATCTA 810

RESULT 13
US-09-949-016-4828
; Sequence 4828, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4828
; LENGTH: 3416
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-4828

Query Match          39.4%; Score 25.2; DB 3; Length 3416;
Best Local Similarity 71.7%; Pred. No. 26;
Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 GATCCGGAAGAGGTGTTCAAGCGCTGGAGAGCTGTTAGCAA 46
DB 765 GAACCGTCTGAGCTGATCAATGCCAGAGAGCTGTATCTA 810

RESULT 14
US-09-949-016-4829
; Sequence 4829, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
```

SEQ ID NO 4829
LENGTH: 3490
TYPE: DNA
ORGANISM: Human
US-09-949-016-4829

Query Match 39.4%; Score 25.2; DB 3; Length 3490;
Best Local Similarity 71.7%; Pred. No. 26;
Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 GATCCGCAAGAGGTGTTCAACGCTGGAGAGAGCTGTTAGCAA 46
DB 765 GAACCTTCTGGAGCTGATCAATGCCAGAGAGAGCTGTATCTA 810

RESULT 15
US-09-949-016-4830
Sequence 4830; Application US/09949016
Patent No. 6812339

GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4830
LENGTH: 3490
TYPE: DNA
ORGANISM: Human
US-09-949-016-4830

Query Match 39.4%; Score 25.2; DB 3; Length 3490;
Best Local Similarity 71.7%; Pred. No. 26;
Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 GATCCGCAAGAGGTGTTCAACGCTGGAGAGAGCTGTTAGCAA 46
DB 765 GAACCTTCTGGAGCTGATCAATGCCAGAGAGAGCTGTATCTA 810

Search completed: February 27, 2006, 06:46:23
Job time: 17.2601 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 27, 2006, 06:33:05 ; Search time 351.203 Seconds
(without alignments)
9873.061 Million cell updates/sec

Title: US-10-789-164-4

Perfect score: 61
Sequence: 1 gatccaagtcgaagaactg.....aagctgcctcaagaactgtca 61

Scoring table: IDENTITY NUC
Gapop 10_0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba: *
2: gb_in: *
3: gb_env: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pr: *
9: gb_ro: *
10: gb_sci: *
11: gb_sy: *
12: gb_un: *
13: gb_vl: *
14: gb_hcg: *
15: gb_pl: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	31.6	51.8	235519	14	AC164258
C 2	31.6	51.8	243489	14	AC164258 Bos tauru
C 3	31.4	51.5	243489	14	AC164258 Bos tauru
C 4	30.8	50.5	36748	12	AF043703
C 5	30.8	50.5	260699	14	AC006893
C 6	30.4	49.8	110000	13	AY653733-06
C 7	30.2	49.5	549	6	AX654322
C 8	30.2	49.5	110000	15	AK107375
C 9	30.2	49.5	110000	15	AK107375 Oryza sat
C 10	30.2	49.5	126938	15	AP008213-061
C 11	30.2	49.5	126938	15	AP008213 Oryza sat
C 12	29.8	48.9	414	2	PFALL1ER
C 13	29.8	48.9	414	2	PFALL1ER
C 14	29.8	48.9	414	2	PFALL1ER
C 15	29.8	48.9	414	2	PFALL1ER
C 16	29.8	48.9	414	2	PFALL1ER
C 17	29.8	48.9	414	2	PFALL1ER
C 18	29.8	48.9	414	2	PFALL1ER

19	29.8	48.9	6409	2	PFAPMMMSA
20	29.6	48.5	1280	15	AK061230
21	29.6	48.5	1741	15	AK071110
22	29.6	48.5	12402	1	AE013431
23	29.6	48.5	110000	15	AP008216-196
24	29.6	48.5	135789	15	AC051634
25	29.6	48.5	300029	15	AE017113
26	29.4	48.2	607	15	AY096696
27	29.4	48.2	843	15	AB035444
28	29.4	48.2	874	15	AY066483
29	29.4	48.2	900	15	AY063997
30	29.4	48.2	107200	15	AC006551
31	29.4	48.2	170120	14	AC145728
32	29.2	47.9	38138	2	AF003148
33	29.2	47.5	330724	14	CEY67H2
34	28.8	47.2	2145	2	DIHSC90R
35	28.8	47.2	3720	13	S76368
36	28.8	47.2	43658	13	HSV3P68N
37	28.8	47.2	112930	13	HSV3P68N
38	28.8	47.2	119958	14	AY914084
39	28.8	47.2	224093	14	AC156886
40	28.6	46.9	3039	5	CR942671
41	28.6	46.9	3042	5	BC060352
42	28.6	46.9	110000	1	CR848038-01
43	28.4	46.6	354	6	AR550464
44	28.4	46.6	476	15	AF317661
45	28.4	46.6	31853	2	CER44D12

ALIGNMENTS

RESULT 1
AC164258/c
LOCUS
DEFINITION
Bos taurus clone CH240-148E22, *** SEQUENCING IN PROGRESS ***, 28
unordered pieces.

ACCESSION
AC164258.2 GI:68300969
VERSION
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
KEYWORDS
Bos taurus (cow)
SOURCE
Bos taurus
ORGANISM

REFERENCE

AUTHORS

Muzny, D., Maric, M., Metzger, M., Lee, A., Adams, C., Alder, J., Allen, C., Allen, R., Albrooks, S., Amin, A., Anguiano, D., Anylebech, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Bismail, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Bunay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Genter, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Cleveland, C., Cockrell, R., Cox, C., Coyler, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dint, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, T., Foster, P., Fraser, C., Gabisel, A., Garcia, R., Garcia, J., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvery, P., Havlak, P., Hawes, A., Henderson, N., Hernandez, R., Hines, S., Hladik, S., Hodge, A., Hogue, M., Hollins, B., Howells, S., Hulik, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpich, S., Kelly, S., Khan, Z., King, L., Kovar, C., Krawiec, C., Kraft, C., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenschewski, L., Louised, H., Lozano, R., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Mallory, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,

Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
 Narkervic, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Parks, K.,
 Nwokedi, O., Okwou, G., Olarunsa, A., Pal, S., Parks, K.,
 Paaterak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
 Plapper, F., Poldinger, A., Popovic, D., Primus, E., Pu, L.,
 Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
 Rell, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rivers, C., Rodkey, T., Rojas, A., Rose, R., Ruiz, S.,
 Sanders, W., Savary, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajic, D.,
 Sneed, A., Sodergren, E., Song, X.-Z., Sorrell, R., Sosa, J.,
 Staimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umanal, K.,
 Valas, R., Vera, V., Villanueva, D., Waldron, L., Walker, B., Wang, J.,
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
 Williams, G., Willson, R., Wleczek, R., Woodson, H., Worley, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
 Niederhausern, A., Weise, R., Smith, D.R., Holt, R.A., Smith, H.O.,
 Weinstock, G., and Gibbs, R.A.
 Direct Submission
 Unpublished
 2 (bases 1 to 235519)
 Worley, K.C.
 Direct Submission
 Submitted (19-JUN-2005) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 235519)
 Cow Genome Sequencing Consortium.
 Direct Submission
 Submitted (01-JUN-2005) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jun 29, 2005 this sequence version replaced gi:167972689.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: FI0D
 Center clone name: CH240-148E22
 ----- Summary Statistics
 Assembly program: Atlas 3.0
 Consensus quality: 221304 bases at least Q40
 Consensus quality: 224455 bases at least Q30
 Consensus quality: 226977 bases at least Q20
 Estimated insert size: 224509 sum-of-contigs estimation
 Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 28 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence

as soon as it is available and the accession number will
 be preserved.

FEATURES	source
1	1231: contig of 1231 bp in length
1232	2496: gap of 1265 bp
2497	15849: contig of 1335 bp in length
15850	18899: gap of 50 bp
15900	33519: contig of 17620 bp in length
33520	33569: gap of 50 bp
33570	47007: contig of 13438 bp in length
47008	47057: gap of 50 bp
47058	49714: contig of 2657 bp in length
49715	49764: gap of 50 bp
49765	56220: contig of 6456 bp in length
56221	56643: gap of 423 bp
56644	58196: contig of 1553 bp in length
58197	58296: gap of unknown length
58297	78107: contig of 19811 bp in length
78108	78157: gap of 50 bp
78158	101919: contig of 23762 bp in length
101920	101969: gap of 50 bp
101970	105635: contig of 3666 bp in length
105636	106018: gap of 383 bp
106019	107358: contig of 1340 bp in length
107359	107930: gap of 572 bp
107931	113033: contig of 5103 bp in length
113034	113097: gap of 64 bp
113098	116951: contig of 23854 bp in length
116952	137051: gap of unknown length
117052	139311: contig of 2260 bp in length
139312	139361: gap of 50 bp
145341	145341: contig of 5980 bp in length
145342	146014: gap of 673 bp
146015	186956: contig of 40942 bp in length
186957	187056: gap of unknown length
187057	189448: contig of 2392 bp in length
189449	189627: gap of 179 bp
189628	200786: contig of 1159 bp in length
200787	200886: gap of unknown length
200887	203015: contig of 2129 bp in length
203016	203490: gap of 475 bp
203491	209254: contig of 5764 bp in length
209255	209354: gap of unknown length
209355	210664: contig of 1210 bp in length
210665	210664: gap of unknown length
210666	211803: contig of 1139 bp in length
211804	211903: gap of unknown length
211904	213042: contig of 1139 bp in length
213043	213142: gap of unknown length
213143	214421: contig of 1279 bp in length
214422	214521: gap of unknown length
214522	216214: gap of 1693 bp in length
216215	216314: gap of unknown length
216316	217352: contig of 1038 bp in length
217353	217452: gap of unknown length
217454	219446: contig of 1994 bp in length
219447	219546: gap of unknown length
219547	235519: contig of 15973 bp in length.

Location/Qualifiers
 1. 235519
 /organism="Bos taurus"
 /mol_type="genomic DNA"
 /db_xref="taxon:9913"
 /clone="CH240-148E22"
 1232. 2496
 /estimated_length=1265
 15850. 15899
 /estimated_length=50
 33520. 33569
 /estimated_length=50
 47008. 47057
 /estimated_length=50
 49715. 49764
 /estimated_length=50

```

gap          gap      56221..56643  

              /estimated_length=423  

gap          gap      58197..58296  

              /estimated_length=unknown  

gap          gap      78108..78157  

              /estimated_length=50  

gap          gap      101920..101969  

              /estimated_length=50  

gap          gap      105636..106018  

              /estimated_length=383  

gap          gap      107359..107930  

              /estimated_length=572  

gap          gap      113034..113097  

              /estimated_length=64  

gap          gap      136952..137051  

Query Match   51.8%; Score 31.6; DB 14; Length 235519;  

Best Local Similarity 80.4%; Pred. No. 39;  

Matches 37; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  

Db            103187 GAAGAAACTCAAGAAGAACGCTGATCAGAAGCTGCCAAGAAA 103142  

Qy           11 GAGAAACCTGCTCAAGAAACCCTGCTCAAGAAAGTGTCTCAAGAAA 56  

             ||||| | | | | | | | | | | | | | | |  

DB         103187 GAAGAAACTCAAGAAGAACGCTGATCAGAAGCTGCCAAGAAA 103142  

RESULT 2  

AC162622  

LOCUS       AC162622  

DEFINITION Bos taurus clone CH240-125A15, *** SEQUENCING IN PROGRESS ***, 31  

ACCESSION  AC162622  

VERSION    AC162622.2 GI:68301295  

KEYWORDS   HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.  

SOURCE     Bos taurus (cow)  

ORGANISM   Bos taurus  

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  

Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;  

Bovidae; Bovinae; Bovineae; Bos.  

1 (bases 1 to 243489)  

Muzny,D.Marie., Metzker,M.Lee., Adrianzon,S., Adams,C., Alder,J.,

```

Allen, C., Allen, H., Alstrooker, S., Amth, A., Anguiano, B., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Baldrastnake, D., Barber, M., Barnstead, M., Bernhardt, F., Bismalo, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carrroll, L., De Andrade, C., Dederich, D., Delgado, O., Denison, S., Detramo, C., Ding, Y., Dinh, H., Dwyer, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durkin, K., Duvall, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gaisels, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gubregorsky, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamli, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hunne, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolyet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kreft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenzenhwa, L., Loulsegged, H., Lozano, R. J., Lu, X., Ma, J., Maneswarar, M., Mahindratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mathiney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Mlilobajivsic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Mundaaes, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nweokemah, O., Okwumodu, G., Olarunmasagoun, A., Pal, S., Parks, K., Paternack, S., Paul, H., Perez, A., Perez, L., Pfannkoch, C., Plopper, F., Poldner, A., Popovic, D., Primus, E., Fu, L., -L.,

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

TITLE
JOURNAL
AUTHORS
REFERENCE

JOURNAL
AUTHORS
REFERENCE

COMMENT

Pullay B., Quirroz J., Rachlin E., Reeves K., Regier M.A., Reign R.,
Relloso B., Relly M., Ren Y., Reuter M., Richards S., Riggs P.,
Rivers C., Rodkey T., Rojas A., Rose M., Rose R., Ruiz S.J.,
Sanders W., Savery G., Scherer S., Scott G., Shatsman S., Shen H.,
Shetty J., Shvartsbeyn A., Slason I., Sitter C.D., Smajs D.,
Sneed A., Sodergren E., Song X.-Z., Sorrelle R., Soosa J.,
Steinle M., Strong R., Sutton A., Swatek A., Tabori P., Taylor C.,
Taylor T., Thomas N., Thomas S., Tingey A., Trejos Z., Uমানan K.,
Valsar Q., Varg S., Villasana D., Waldron L., Walker B., Wang J.,
Wang O., Wang S., Warren C., Warren R., Wei X., White F.,
Williams G., Willson R., Wleczyk R., Wooden H., Worley K.,
Wright D., Wright R., Wu J., Yakub S., Yen J., Yoon L., Yoon V.,
Yu R., Zhang J., Zhou J., Zhou X., Zhao S., Dunn D., von
Niederhausern A., Weise R., Smith D.R., Holt R.A., Smith H.O.,
Weinstock G. and Gibbs R.A.

Unpublished
Direct Submission
2 (bases 1 to 243489)

Morley K.C.
Submitted (31-MAY-2005) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 243489)
Cow Genome Sequencing Consortium.

Unpublished
Direct Submission
Submitted (01-JUN-2005) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

On Jun 29, 2005 this sequence version replaced gi:66796652.
This assembly is a combination of PAC based reads

```

individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

-----Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: FHOI
Center clone name: CH240-125A15

-----Summary Statistics
Assembly program: Atlas 3.0.
Consensus quality: 230956 bases at least Q40
Consensus quality: 234288 bases at least Q30
Consensus quality: 236517 bases at least Q20
Estimated insert size: 234246; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbankdraft\_data.html)
* NOTE: This is a 'working draft' sequence. It currently
* consists of 31 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
11092: contig of 11092 bp in length
11093 11142: gap of 50 bp
11143 12663: contig of 1521 bp in length
12664 12763: gap of unknown length
12764 25285: contig of 12522 bp in length

```

25286	25335	gap of 50 bp
25336	32236	contig of 7901 bp in length
25337	33286	gap of 50 bp
33287	34642	contig of 1356 bp in length
3463	35251	gap of 609 bp
35252	37801	contig of 2550 bp in length
37802	38123	gap of 322 bp
38124	55860	contig of 17737 bp in length
55861	55960	gap of unknown length
55961	68473	contig of 13513 bp in length
68474	69524	gap of 50 bp
69524	74879	contig of 5355 bp in length
74879	75341	gap of 462 bp
75341	77511	contig of 2071 bp in length
77511	77511	gap of unknown length
77512	104543	contig of 27032 bp in length
104544	104931	gap of 388 bp
104932	107567	contig of 2636 bp in length
107568	107617	gap of 50 bp
107618	121698	contig of 14081 bp in length
121699	121748	gap of 50 bp
121749	133466	contig of 1718 bp in length
133467	133566	gap of unknown length
133567	138859	contig of 223 bp in length
138860	135970	gap of 50 bp
135971	159070	contig of 23161 bp in length
159071	159120	gap of 50 bp
159121	1762278	contig of 17158 bp in length
176229	176328	gap of 50 bp
176329	176842	contig of 3514 bp in length
176843	180440	gap of 598 bp
180441	209534	contig of 29094 bp in length
209535	209584	gap of 50 bp
209585	214425	contig of 4841 bp in length
214426	214475	gap of 50 bp
214476	222990	contig of 8515 bp in length
222991	223040	gap of 50 bp
223041	228305	contig of 5525 bp in length
228306	228435	gap of 130 bp
228436	231133	contig of 3688 bp in length
231134	232233	gap of unknown length
232234	232380	contig of 1047 bp in length
232381	233380	gap of unknown length
233381	233491	contig of 1011 bp in length
234492	235491	gap of unknown length
235492	235521	contig of 1030 bp in length
235522	235621	gap of unknown length
235622	235678	contig of 1057 bp in length
236679	236778	gap of unknown length
236779	238364	contig of 1566 bp in length
238365	238464	gap of unknown length
238465	239925	contig of 1461 bp in length
239926	240025	gap of unknown length
240026	241242	contig of 1217 bp in length
241243	241342	gap of unknown length
241343	243489	contig of 2147 bp in length

URES	location/qualifiers
source	1. .243489
	/organism="Bos taurus"
	/mol_type="genomic DNA"
	/db_xref="taxon:9913"
	/clone="CH240-125A15"
gap	11093. .11142
	/estimated_length=50
gap	12664. .12763
	/estimated_length=unknown
gap	25286. .25335
	/estimated_length=50
gap	33337. .33286
	/estimated_length=50
gap	34643. .35251
	/estimated_length=609
gap	37802. .38123

gap	/estimated_length=322
	58661. .55960
	/estimated_length=unknown
gap	69474. .65523
	/estimated_length=50
gap	74879. .75340
	/estimated_length=462
gap	77412. .77511

Query Match	51.8%;	Score 31.6;	DB 14;	Length 243489;
Best Local Similarity	80.4%;	Pred. No. 39;		
Matches 37;	Conservative	0;	Mismatches 9;	Indels 0;
				Gaps 0;

Qy	11	GAAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAGAA	56
Db	37145	GAAGAACTCAACAGAGAAAGTGTATCAAGAGCTGGCCCAAGAA	37190

RESULT 3	LOCUS	DEFINITION
AF317660	441 bp DNA	linear
AF317660	Candida albicans 60S acidic ribosomal protein type P1-B (p1B) gene.	PLN 20-NOV-2000

ACCESSION	AF317660
VERSION	AF317660.1
	GI:11229037

SOURCE	Candida albicans
Oral cavity	100%
Vagina	100%
Intestine	100%
Skin	100%
Urogenital tract	100%
Respiratory tract	100%
Genital tract	100%
Uterus	100%
Placenta	100%
Fetus	100%
Amniotic fluid	100%
Peritoneal fluid	100%
Pericardial fluid	100%
Joint fluid	100%
Eye	100%
Ear	100%
Nose	100%
Throat	100%
Saliva	100%
Sweat	100%
Urine	100%
Feces	100%
Stool	100%
Uterine secretions	100%
Vaginal secretions	100%
Genital secretions	100%
Urogenital secretions	100%
Respiratory secretions	100%
Genital secretions	100%
Uterine secretions	100%
Placental secretions	100%
Fetal secretions	100%
Amniotic secretions	100%
Peritoneal secretions	100%
Pericardial secretions	100%
Joint secretions	100%
Eye secretions	100%
Ear secretions	100%
Nose secretions	100%
Throat secretions	100%
Saliva secretions	100%
Sweat secretions	100%
Urine secretions	100%
Feces secretions	100%
Stool secretions	100%
Uterine secretions	100%
Vaginal secretions	100%
Genital secretions	100%
Urogenital secretions	100%
Respiratory secretions	100%
Genital secretions	100%
Uterine secretions	100%
Placental secretions	100%
Fetal secretions	100%
Amniotic secretions	100%
Peritoneal secretions	100%
Pericardial secretions	100%
Joint secretions	100%
Eye secretions	100%
Ear secretions	100%
Nose secretions	100%
Throat secretions	100%
Saliva secretions	100%
Sweat secretions	100%
Urine secretions	100%
Feces secretions	100%
Stool secretions	100%
Uterine secretions	100%
Vaginal secretions	100%
Genital secretions	100%
Urogenital secretions	100%
Respiratory secretions	100%
Genital secretions	100%
Uterine secretions	100%
Placental secretions	100%
Fetal secretions	100%
Amniotic secretions	100%
Peritoneal secretions	100%
Pericardial secretions	100%
Joint secretions	100%
Eye secretions	100%
Ear secretions	100%
Nose secretions	100%
Throat secretions	100%
Saliva secretions	100%
Sweat secretions	100%
Urine secretions	100%
Feces secretions	100%
Stool secretions	100%
Uterine secretions	100%
Vaginal secretions	100%
Genital secretions	100%
Urogenital secretions	100%
Respiratory secretions	100%
Genital secretions	100%
Uterine secretions	100%
Placental secretions	100%
Fetal secretions	100%
Amniotic secretions	100%
Peritoneal secretions	100%
Pericardial secretions	100%
Joint secretions	100%
Eye secretions	100%
Ear secretions	100%
Nose secretions	100%
Throat secretions	100%
Saliva secretions	100%
Sweat secretions	100%
Urine secretions	100%
Feces secretions	100%
Stool secretions	100%
Uterine secretions	100%
Vaginal secretions	100%
Genital secretions	100%
Urogenital secretions	100%
Respiratory secretions	100%
Genital secretions	100%
Uterine secretions	100%
Placental secretions	100%
Fetal secretions	100%
Amniotic secretions	100%
Peritoneal secretions	100%
Pericardial secretions	100%
Joint secretions	100%
Eye secretions	100%
Ear secretions	100%
Nose secretions	100%
Throat secretions	100%
Saliva secretions	100%
Sweat secretions	100%
Urine secretions	100%
Feces secretions	100%
Stool secretions	100%
Uterine secretions	100%
Vaginal secretions	100%
Genital secretions	100%
Urogenital secretions	100%
Respiratory secretions	100%
Genital secretions	100%
Uterine secretions	100%
Placental secretions	100%
Fetal secretions	100%
Amniotic secretions	100%
Peritoneal secretions	100%
Pericardial secretions	100%
Joint secretions	100%
Eye secretions	100%
Ear secretions	100%
Nose secretions	100%
Throat secretions	100%
Saliva secretions	100%
Sweat secretions	100%
Urine secretions	100%
Feces secretions	100%
Stool secretions	100%
Uterine secretions	100%
Vaginal secretions	100

Classification: Eukaryota; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; mitosporic Saccharomycetales; Candida.

REFERENCE
1. (Dabas 1 CO 711)
AUTHORS
Abramczyk, D., Tchorzewski, M. and Grankowski, N.

TITLE Cloning, expression and purification of the acidic ribosomal protein from *Candida albicans*

JOURNAL Unpublished

REFERENCE
2 (bases 1 to 441)
Abramczyk, D., Tchorzewski, M. and Grankowski, N

Submitted (01-NOV-2000) Department of Molecular Biology, Maria Curie-Skłodowska University, Akademicka 19, Lublin 20-033, Poland

FEATURES	LOCATION/QUALITIES
SOURCE	1. .441

```
/organism="Candida albicans"  
/mol_type="genomic DNA"
```

```

/db xref="taxon:5476"
/sciain="CCM 8215"

```

```

gene
  <0.12534
  /gene="p1B"

```

```
/gene="p1B"
```

CDS	/P-0000-000 80000000
	8 . . 334

```
/gene= p1b
/note="type P1"
```

```

/conn_buffers=1
/transl_table=12

```

```

/product= 000 uc1a1c
/protein_id="AAG33241

```

```
/usr/bin/perl -s.11220000
/translation="MSTEASV
```

FAVVEGUEEEDS ST. 2000
GFGLED"

ORIGIN

Query Match	51.5%;	score 31
Best Local Similarity	77.6%;	Pred. Nc

Machine	30, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000	Machine
Machine	30, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000	Machine

[illegible]

DB 234 GAAAGCAGCAGCGAAGGACCCGAGAAC

LOCUS	AF043703	36748 bp	DNA	linear	INV 27-MAY-2005				
DEFINITION	Caenorhabditis elegans cosmid W10C8, complete sequence.								
ACCESSION	AF043703								
VERSION	AF043703.2	GI:13384574							
KEYWORDS	HTG.								
SOURCE	Caenorhabditis elegans								
ORGANISM	Eukaryote; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.								
REFERENCE	1 (bases 1 to 36748)								
AUTHORS	C. elegans Sequencing Consortium								
CONSTRM	Genome sequence of the nematode <i>C. elegans</i> : a platform for investigating biology								
TITLE	Science 282 (5396), 2012-2018 (1998)								
JOURNAL	9851916								
REFERENCE	2 (bases 1 to 36748)								
AUTHORS	Jones, K., Tin-Mollam, A. and Blair, T.								
TITLE	The sequence of <i>C. elegans</i> cosmid W10C8								
JOURNAL	Unpublished (2001)								
REFERENCE	3 (bases 1 to 36748)								
AUTHORS	Waterston, R.								
TITLE	Direct Submission								
JOURNAL	Submitted (20-JUN-1998) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA								
REFERENCE	4 (bases 1 to 36748)								
AUTHORS	Waterston, R.								
TITLE	Direct Submission								
JOURNAL	Submitted (04-AUG-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA								
REFERENCE	5 (bases 1 to 36748)								
AUTHORS	Waterston, R.								
TITLE	Direct Submission								
JOURNAL	Submitted (20-MAR-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA								
REFERENCE	6 (bases 1 to 36748)								
AUTHORS	Waterston, R.								
TITLE	Direct Submission								
JOURNAL	Submitted (05-OCT-2001) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA								
REFERENCE	7 (bases 1 to 36748)								
AUTHORS	Waterston, R.								
TITLE	Direct Submission								
JOURNAL	Submitted (04-MAY-2002) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA								
REFERENCE	8 (bases 1 to 36748)								
AUTHORS	Waterston, R.								
TITLE	Direct Submission								
JOURNAL	Submitted (21-NOV-2002) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA								
REFERENCE	9 (bases 1 to 36748)								
AUTHORS	Waterston, R.								
TITLE	Direct Submission								
JOURNAL	Submitted (09-DEC-2002) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA								
REFERENCE	10 (bases 1 to 36748)								
AUTHORS	Waterston, R.								
TITLE	Direct Submission								
JOURNAL	Submitted (10-FEB-2003) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA								
REFERENCE	11 (bases 1 to 36748)								
AUTHORS	Waterston, R.								
TITLE	Direct Submission								
JOURNAL	Submitted (10-MAR-2003) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA								

REFERENCE	12 (bases 1 to 36748)
AUTHORS	Wilson, R.
TITLE	Direct SubMISSION
JOURNAL	Submitted (03-DEC-2003) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA
REFERENCE	13 (bases 1 to 36748)
AUTHORS	
CONSRMT	WormBase Consortium
TITLE	Direct SubMISSION
JOURNAL	Submitted (22-SEP-2004) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA
REFERENCE	14 (bases 1 to 36748)
AUTHORS	
CONSRMT	WormBase Consortium
TITLE	Direct SubMISSION
JOURNAL	Submitted (27-MAY-2005) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA
COMMENT	On Mar 20, 2001 this sequence version replaced gi:2804482. Contained in:

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality ≥ 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

For a graphical representation of this clone sequence and its analysis see:
<http://www.wormbase.org/db/seq/sequence?name=W10C8;class=Sequence>

NEIGHBORING CLONE INFORMATION

The 5' clone is W03D8, 2500 bp overlap; the 3' clone is Y71F9AL, 200 bp overlap. Actual start of this clone is at base position 1 of W10C8; actual end is at 36748 of W10C8.

NOTES

coding sequences below are the result of integration and manual review of the following data : computer analysis using the program GeneIndex (P. Green and L. Hillier, personal communication), the large scale EST project of Yuiji Kohara (http://www.ddbj.nig.ac.jp/c-elegans/hml/CE_INDEX.html) and The C. elegans ORFome cloning project (<http://worldf.dcel.harvard.edu/>), similarity to other proteins from Blastx analyses (<http://blast.wustl.edu/>), sequence conservation with C. briggsae using Jij Kent's WABA alignment program (Genome Research 10:1115-1125, 2000), individual C. elegans GenBank submissions, and personal communications with C. elegans researchers. tRNAs are predicted using the program tRNAscan-SE (Lowie, T.M. and Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

FEATURES

```
source
1.36748
/organism="Caenorhabditis elegans"
/mol_type="genomic DNA"
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="I"
```

gene
/clone="W10C8"
complement(56..6878)
/gene="pop-1"
/locus_tag="W10C8.2"
complement(join(56..525,1465..2017,5068..5186,6707..6878))
/gene="pop-1"
/locus_tag="W10C8.2"
/standard_name="W10C8.2"
/note="contains similarity to Pfam domain PF00505 (HMG
high mobility group) box); coded for by the following C.
elegans cDNAs: CKS85925, CKS87970, OSTF153D9_1,
OSTR153D9_1, U37532, OSTF30017H05"
/codon_start=1
/product="posterior pharynx defect protein 1"
/protein_id="AAK21504.1"
/db_xref="GI:13384579"
/db_xref="Wormbase:W10C8.2"
/translation="MADELDGVKVFRRDEADDDPMISGETSEQOLADKKEAVNE
AELDAGRNPSIDVLKSAFPEVPMSPSPGLMHSFSGYSAALPMFMPLPMNPVYA
ALRSPILMEPMGASPTPMPPPSPVYGAALAAAKOHFENMAPIIMRAGHPNMONG
MPYPMHPSNAPQVNDRAQCGGAKKDDHYKPLINAFMFMKERNKALLBEIGNEK
QSAELNLELGRHMDLSEQAKYFEWAKKDETHKERYPMWASENAVNKKTKR
RDKSIPSENNQKCRAPFGVNTMEMCKFCRKKECYATDRSGSDITVDSQGRGT
SGAYSSSSSPSPKANAGIALTQQOQAAAHMTLMQMRLGTTGASTGATVPSPLASS
AGRSPIDANASDSVDVEEDQIDPTVMQOTHTDMLMQESMCTI"
complement(9460..9841)
/locus_tag="W10C8.3"
complement(join(9460..9561,9737..9841))
/locus_tag="W10C8.3"
/standard_name="W10C8.3"
/note="contains similarity to Hepatitis B virus SHBs;
TR:091164"
/codon_start=1
/product="Hypothetical protein W10C8.3"
/protein_id="AAK21501.2"
/db_xref="GI:26251556"
/db_xref="Wormbase:W10C8.3"
/translation="MESADSWLSPNPQIRQARKMDEFSRCSLRATIALIVCYAW
LNNTCLSFNSFRNFLLPQPL"
23308..26380
/gene="Ccb-2"
/locus_tag="W10C8.1"
join(23308..23373,23547..23642,23698..23820,24632..25174,
25843..25974,26267..26380)
/gene="Ccb-2"
/locus_tag="W10C8.1"
/standard_name="W10C8.1"
/note="contains similarity to Pfam domains PF00774
(Dihydropyridine sensitive L-type calcium channel (Beta
subunit)), PF00018 (SH3 domain); coded for by the
following C. elegans cDNAs: OSTF081A3_1, OSTT081A3_1"
/codon_start=1
/product="Calcium channel, beta subunit protein 2"
/protein_id="AAK21500.1"
/db_xref="GI:13384575"
/db_xref="Wormbase:W10C8.1"
Query Match 50.5%; Score 30.8; DB 2; Length 36748;
Beet Local Similarity 70.7%; Pred. No. 78;
Matches 41; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
OY 1 GATCGAATGGAAGAACTGCTCAAGAAACCGCTCAAGAACTGCTCAAGAACT 58
DB 31755 GAGCCAAAGTCAAGTCGCTCAAGAAAGCATCTCAAGAGCGTCTCGAAGACT 31698
RESULT 5 260699 bp DNA linear HTG 26-FEB-1999
LOCUS AC006893
DEFINITION Caenorhabditis elegans clone Y71f9, *** SEQUENCING IN PROGRESS ***,
43 unordereed pieces.
ACCESSION AC006893
VERSION AC006893.2 GI:4309906

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
HTG; HTGS PHASE1.
Caenorhabditis elegans
Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 260699)
Waterston,R.H.
The sequence of Caenorhabditis elegans clone
Unpublished
2 (bases 1 to 260699)
Waterston,R.H.
Direct Submission
Submitted (24-FEB-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Mar 1, 1999 this sequence version replaced gi:4263455.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 43 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 3303: contig of 3303 bp in length
3304 3312: gap of unknown length
3313 6195: contig of 2883 bp in length
6196 6204: gap of unknown length
6205 8705: contig of 2501 bp in length
8706 8714: gap of unknown length
8715 27303: contig of 18589 bp in length
27304 27312: gap of unknown length
27313 29898: contig of 2586 bp in length
29899 29907: gap of unknown length
29908 32748: contig of 2841 bp in length
32749 32757: gap of unknown length
32758 42435: contig of 9678 bp in length
42436 42444: gap of unknown length
42445 45447: contig of 3003 bp in length
45448 45456: gap of unknown length
45457 56578: contig of 11122 bp in length
56579 56587: gap of unknown length
56588 59294: contig of 2707 bp in length
59295 59303: gap of unknown length
59304 70657: contig of 11353 bp in length
70658 70665: gap of unknown length
70666 78234: contig of 7569 bp in length
78235 78243: gap of unknown length
78244 81287: contig of 3043 bp in length
81288 87906: contig of 6611 bp in length
87907 87915: gap of unknown length
87916 90315: contig of 2400 bp in length
90316 90324: gap of unknown length
90325 93415: contig of 3091 bp in length
93416 93424: gap of unknown length
93425 96332: contig of 2908 bp in length
96333 96341: gap of unknown length
96342 102339: contig of 5998 bp in length
102340 102348: gap of unknown length
102349 104720: contig of 2372 bp in length
104721 104729: gap of unknown length
104730 113259: contig of 8530 bp in length
113260 113268: gap of unknown length
113269 117369: contig of 4101 bp in length
117370 117378: gap of unknown length
117379 119526: contig of 2148 bp in length
119527 119535: gap of unknown length
119536 121798: contig of 2263 bp in length
121799 121807: gap of unknown length
121808 130291: contig of 8484 bp in length
130292 130300: gap of unknown length
130301 146526: contig of 16226 bp in length

* 146527 146535: gap of unknown length
* 146536 156661: contig of 10126 bp in length
* 156662 156670: gap of unknown length
* 156671 161833: contig of 5163 bp in length
* 161834 161842: gap of unknown length
* 161843 169581: contig of 7739 bp in length
* 169582 169590: gap of unknown length
* 169591 174622: contig of 5032 bp in length
* 174623 174631: gap of unknown length
* 174632 177362: contig of 2731 bp in length
* 177363 177371: gap of unknown length
* 177372 180479: contig of 3108 bp in length
* 180480 180488: gap of unknown length
* 180489 185112: contig of 4824 bp in length
* 185113 185321: gap of unknown length
* 185322 189706: contig of 4385 bp in length
* 189707 189715: gap of unknown length
* 189716 199829: contig of 10214 bp in length
* 199930 199938: gap of unknown length
* 199939 209300: contig of 9362 bp in length
* 209301 209309: gap of unknown length
* 209310 215261: contig of 5952 bp in length
* 215262 215269: gap of unknown length
* 215270 223804: contig of 8535 bp in length
* 223805 223812: gap of unknown length
* 223813 225776: contig of 1964 bp in length
* 225777 225784: gap of unknown length
* 225785 229087: contig of 3303 bp in length
* 229088 229095: gap of unknown length
* 229096 241882: contig of 12787 bp in length
* 241883 241890: gap of unknown length
* 241891 249775: contig of 7885 bp in length
* 249776 249783: gap of unknown length
* 249784 253545: contig of 3762 bp in length
* 253546 253553: gap of unknown length
* 253554 260699: contig of 7146 bp in length.

FEATURES
source 1. 260699
Location/Qualifiers
/organism="Caenorhabditis elegans"
/mol_type="genomic DNA"
/db_xref="taxon:6239"
/clone="Y71F9"
3304..3312
/estimated_length=unknown
6196..6204
/estimated_length=unknown
8706..8714
/estimated_length=unknown
27304..27312
/estimated_length=unknown
29899..29907
/estimated_length=unknown
32749..32757
/estimated_length=unknown
42436..42444
/estimated_length=unknown
45448..45456
/estimated_length=unknown
55579..55587
/estimated_length=unknown
59295..59303
/estimated_length=unknown
70657..70665
/estimated_length=unknown
78235..78243
/estimated_length=unknown
81287..81295
/estimated_length=unknown
87907..87915
/estimated_length=unknown
90316..90324
/estimated_length=unknown
93416..93424

gap /estimated_length=unknown
96333..96341
/estimated_length=unknown
102340..102348
gap /estimated_length=unknown
104721..104729
gap /estimated_length=unknown
113260..113268
gap /estimated_length=unknown
117370..117378
gap /estimated_length=unknown
119527..119535
gap /estimated_length=unknown
121799..121807
gap /estimated_length=unknown
130292..130300
gap /estimated_length=unknown
146527..146535
gap /estimated_length=unknown
156662..156670
gap /estimated_length=unknown
161834..161842
gap /estimated_length=unknown
169582..169590
gap /estimated_length=unknown
174623..174631
gap /estimated_length=unknown
177363..177371
gap /estimated_length=unknown
180480..180488
gap /estimated_length=unknown
185113..185121
gap /estimated_length=unknown
189707..189715
gap /estimated_length=unknown
199930..199938
gap /estimated_length=unknown
209301..209309
gap /estimated_length=unknown
215262..215269
gap /estimated_length=unknown
223805..223812
gap /estimated_length=unknown
225777..225784
gap /estimated_length=unknown
229088..229095
gap /estimated_length=unknown
241883..241890
gap /estimated_length=unknown

Query Match 50.5%; Score 30.8; DB 14; Length 260699;
Best Local Similarity 70.7%; Pred. No. 67;
Matches 41; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 1 GATCCAGTGAAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTCAAGAACT 58
DB 200571 GAGCCAGTGAAGTCTGCTCAAGAACATCTGACCAAGACCTGCTGAGAACT 200628

RESULT 6
AY653733_06
WPCOMMENT

Sequence split into 12 fragments LOCUS AY653733 Accession AY653733
Fragment Name Begin End
AY653733_00 1 110000
AY653733_01 100001 210000
AY653733_02 310000 410000
AY653733_03 300001 410000
AY653733_04 400001 510000
AY653733_05 500001 610000
AY653733_06 600001 710000
AY653733_07 700001 810000
AY653733_08 800001 910000

AY653733.09 900001 1010000
 AY653733.10 1000001 1100000
 AY653733.11 1000001 1181404
 Continuation (7 of 12) of AY653733 from base 600001 (AY653733 Acanthamoeba polyphaga mit

Query Match 49.8%; Score 30.4; DB 13; Length 110000;
 Best Local Similarity 71.4%; Pred. No. 94;
 Matches 40; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Oy 5 CACGTGAGAGAACTCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAGAACTCT 60
 Db 99868 CCAGTTCAAGAACTCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAGAACTCT 99923

RESULT 7
 LOCUS AK654322 549 bp DNA linear PAT 22-MAR-2003
 DEFINITION Sequence 4192 from Patent WO03000898.
 ACCESSION AK654322
 VERSION AK654322.1 GI:29157136
 KEYWORDS

SOURCE
 ORGANISM *Oryza sativa*
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; *Oryza*.

REFERENCE
 AUTHORS Chang, H.S., Chen, W., Cooper, B., Glazebrook, J., Goff, S.A., Hou, Y.M.,
 Katsagiri, F., Quan, S., Tao, Y., Whitman, S., Xie, Z., Zhu, T. and Zou, G.
 TITLE Plant genes involved in defense against pathogens
 JOURNAL Patent: WO 03000898-A 4192 03-JAN-2003;
 Syngenta Participations AG (CH)

FEATURES
 source
 1..549
 /organism="Oryza sativa"
 /mol_type="unassigned DNA"
 /db_xref="taxon:4530"

ORIGIN
 Query Match 49.5%; Score 30.2; DB 6; Length 549;
 Best Local Similarity 74.5%; Pred. No. 1.6e+02;
 Matches 38; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Oy 11 GAGAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAGAACTGTA 61
 Db 205 GATGAACTGTTGAGAGAGCGACCTCTCTGTGAACTATTGAGAGAGCTGGA 255

RESULT 8
 LOCUS AK107375 1188 bp mRNA linear PLN 24-JUL-2003
 DEFINITION *Oryza sativa* (japonica cultivar-group) cDNA clone:002-127-B08, full
 insert sequence.
 ACCESSION AK107375
 VERSION AK107375.1 GI:3292584
 KEYWORDS FLI CDNA; oligo capping.
 SOURCE *Oryza sativa* (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; *Oryza*.

REFERENCE
 AUTHORS 1 The Rice Full-Length cDNA Consortium, National Institute of
 Agrobiological Sciences Rice Full-Length cDNA Project Team;
 Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K.,
 Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H.,
 Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C.,
 Ohneda, K., Shishiki, T., Foundation of Advancement of International
 Science Genome Sequencing & Analysis Group; Otsu, Y., Murakami, K.,
 Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y.,
 Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M.,
 Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Nikura, J.,
 Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Mura, J.,

Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN;
 Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S.,
 Hara, A., Hashidume, W., Hayatsu, N., Imocani, K., Ishii, Y., Itoh, M.,
 Kagawa, I., Kondo, S., Kono, H., Miyazaki, A., Otsu, N., Oka, Y.,
 Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,
 Yoshino, M. and Hayashizaki, Y.
 TITLE Collection, mapping, and annotation of over 28,000 cDNA clones from
 japonica rice
 JOURNAL Science 301 (5631), 376-379 (2003)
 PUBMED 12869764

REFERENCE
 AUTHORS 2 (bases 1 to 1188)
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K.,
 Fujimura, T., Fukuda, S., Hanegaki, T., Hara, A., Hashizume, W.,
 Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hizumoto, K., Hirooka, T.,
 Hori, F., Hotta, I., Iida, Y., Iida, Y., Ikeda, R., Imamura, K.,
 Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I.,
 Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M.,
 Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M.,
 Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,
 Koya, S., Kurihara, C., Kurosaki, T., Kusumegi, T., Li, C., Lu, M.,
 Masuda, H., Matsubara, K., Matsuyama, T., Mura, J., Miyazaki, A.,
 Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M.,
 Namiki, T., Narikawa, R., Nikura, J., Nishi, K., Nomura, K.,
 Numasaki, R., Ohneda, E., Ono, M., Ohtsuki, K., Oka, M., Ooka, H.,
 Otsu, N., Oka, Y., Otsu, Y., Ryu, R., Saitoh, H., Sakai, C., Sakai, K.,
 Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K.,
 Shinagawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S.,
 Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y.,
 Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A.,
 Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W.,
 Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S. and
 Yoshimura, A.
 TITLE Direct Submission
 JOURNAL Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of
 Agrobiological Sciences, Department of Molecular Genetics, Head of
 Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki
 305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp)
 Tel:81-29-838-7007, Fax:81-29-838-7007
 This clone is one of the 28k full-length cDNA clones from japonica
 rice.
 URL: <http://cdna01.dna.affrc.go.jp/cDNA/NIAS>
 NIAS Rice Full-Length cDNA Project Team: Kikuchi, S., Satoh, K.,
 Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J.,
 Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T.,
 Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T. and
 Yamamoto, M.
 FAIS Genome Sequencing & Analysis Group: Otsu, Y., Iida, Y.,
 Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M.,
 Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Mura, J.,
 Mizuno, K., Narikawa, R., Nikura, J., Oka, M., Ryu, R., Sugano, S.,
 Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S.,
 Yoshimura, A., Matsubara, K. and Murakami, K.
 Genome Exploration Research Group in Riken Genomic Sciences Center
 and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K.,
 Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hangaki, T.,
 Hara, A., Hashidume, W., Hayashida, K., Hayatsu, N., Hizumoto, K.,
 Hiraoka, T., Hori, F., Iida, Y., Imamura, K., Imocani, K., Ishii, Y.,
 Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J.,
 Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
 Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ono, M., Otsu, N.,
 Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H.,
 Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,
 Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F.,
 Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K.,
 Yasunishi, A. and Hayashizaki, Y.
 Location/Qualifiers
 1..1188
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultiivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="002-127-B08"

ORIGIN

Query Match 49.5%; Score 30.2; DB 15; Length 1188;
Best Local Similarity 74.5%; Pred. No. 1.5e+02;
Matches 38; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

OY 11 GAGAACTGCTCAAGAAACCGCTGCTCAAGAGCTGCTCAAGAACTGTA 61
DB 392 GATGAAGCTGTGAAGAGGACGCTCTCTGAAGCTATTGAAGAGCTGGA 442

RESULT 9
AP008213_061/c
WPCOMMENT

Sequence split into 297 fragments LOCUS AP008213 Accession AP008213

Fragment Name	Begin	End
AP008213_000	1	110000
AP008213_001	100001	210000
AP008213_002	200001	310000
AP008213_003	300001	410000
AP008213_004	400001	510000
AP008213_005	500001	610000
AP008213_006	600001	710000
AP008213_007	700001	810000
AP008213_008	800001	910000
AP008213_009	900001	1010000
AP008213_010	1000001	1110000
AP008213_011	1100001	1210000
AP008213_012	1200001	1310000
AP008213_013	1300001	1410000
AP008213_014	1400001	1510000
AP008213_015	1500001	1610000
AP008213_016	1600001	1710000
AP008213_017	1700001	1810000
AP008213_018	1800001	1910000
AP008213_019	1900001	2010000
AP008213_020	2000001	2110000
AP008213_021	2100001	2210000
AP008213_022	2200001	2310000
AP008213_023	2300001	2410000
AP008213_024	2400001	2510000
AP008213_025	2500001	2610000
AP008213_026	2600001	2710000
AP008213_027	2700001	2810000
AP008213_028	2800001	2910000
AP008213_029	2900001	3010000
AP008213_030	3000001	3110000
AP008213_031	3100001	3210000
AP008213_032	3200001	3310000
AP008213_033	3300001	3410000
AP008213_034	3400001	3510000
AP008213_035	3500001	3610000
AP008213_036	3600001	3710000
AP008213_037	3700001	3810000
AP008213_038	3800001	3910000
AP008213_039	3900001	4010000
AP008213_040	4000001	4110000
AP008213_041	4100001	4210000
AP008213_042	4200001	4310000
AP008213_043	4300001	4410000
AP008213_044	4400001	4510000
AP008213_045	4500001	4610000
AP008213_046	4600001	4710000
AP008213_047	4700001	4810000
AP008213_048	4800001	4910000
AP008213_049	4900001	5010000
AP008213_050	5000001	5110000
AP008213_051	5100001	5210000
AP008213_052	5200001	5310000
AP008213_053	5300001	5410000
AP008213_054	5400001	5510000
AP008213_055	5500001	5610000
AP008213_056	5600001	5710000

AP008213_057	5700001	5810000
AP008213_058	5800001	5910000
AP008213_059	5900001	6010000
AP008213_060	6000001	6110000
AP008213_061	6100001	6210000
AP008213_062	6200001	6310000
AP008213_063	6300001	6410000
AP008213_064	6400001	6510000
AP008213_065	6500001	6610000
AP008213_066	6600001	6710000
AP008213_067	6700001	6810000
AP008213_068	6800001	6910000
AP008213_069	6900001	7010000
AP008213_070	7000001	7110000
AP008213_071	7100001	7210000
AP008213_072	7200001	7310000
AP008213_073	7300001	7410000
AP008213_074	7400001	7510000
AP008213_075	7500001	7610000
AP008213_076	7600001	7710000
AP008213_077	7700001	7810000
AP008213_078	7800001	7910000
AP008213_079	7900001	8010000
AP008213_080	8000001	8110000
AP008213_081	8100001	8210000
AP008213_082	8200001	8310000
AP008213_083	8300001	8410000
AP008213_084	8400001	8510000
AP008213_085	8500001	8610000
AP008213_086	8600001	8710000
AP008213_087	8700001	8810000
AP008213_088	8800001	8910000
AP008213_089	8900001	9010000
AP008213_090	9000001	9110000
AP008213_091	9100001	9210000
AP008213_092	9200001	9310000
AP008213_093	9300001	9410000
AP008213_094	9400001	9510000
AP008213_095	9500001	9610000
AP008213_096	9600001	9710000
AP008213_097	9700001	9810000
AP008213_098	9800001	9910000
AP008213_099	9900001	10010000
AP008213_100	10000001	10110000
AP008213_101	10100001	10210000
AP008213_102	10200001	10310000
AP008213_103	10300001	10410000
AP008213_104	10400001	10510000
AP008213_105	10500001	10610000
AP008213_106	10600001	10710000
AP008213_107	10700001	10810000
AP008213_108	10800001	10910000
AP008213_109	10900001	11010000
AP008213_110	11000001	11110000
AP008213_111	11100001	11210000
AP008213_112	11200001	11310000
AP008213_113	11300001	11410000
AP008213_114	11400001	11510000
AP008213_115	11500001	11610000
AP008213_116	11600001	11710000
AP008213_117	11700001	11810000
AP008213_118	11800001	11910000
AP008213_119	11900001	12010000
AP008213_120	12000001	12110000
AP008213_121	12100001	12210000
AP008213_122	12200001	12310000
AP008213_123	12300001	12410000
AP008213_124	12400001	12510000
AP008213_125	12500001	12610000
AP008213_126	12600001	12710000
AP008213_127	12700001	12810000
AP008213_128	12800001	12910000
AP008213_129	12900001	13010000

```

AP008213_130 13000001 13110000
AP008213_131 13100001 13210000
AP008213_132 13200001 13310000
AP008213_133 13300001 13410000
AP008213_134 13400001 13510000
AP008213_135 13500001 13610000
AP008213_136 13600001 13710000
AP008213_137 13700001 13810000
AP008213_138 13800001 13910000
AP008213_139 13900001 14010000
AP008213_140 14000001 14110000
AP008213_141 14100001 14210000
AP008213_142 14200001 14310000
AP008213_143 14300001 14410000
AP008213_144 14400001 14510000
AP008213_145 14500001 14610000
AP008213_146 14600001 14710000
AP008213_147 14700001 14810000
AP008213_148 14800001 14910000
AP008213_149 14900001 15010000
AP008213_150 15000001 15110000
AP008213_151 15100001 15210000
AP008213_152 15200001 15310000
AP008213_153 15300001 15410000
AP008213_154 15400001 15510000
AP008213_155 15500001 15610000
AP008213_156 15600001 15710000
AP008213_157 15700001 15810000
AP008213_158 15800001 15910000
AP008213_159 15900001 16010000
AP008213_160 16000001 16110000
AP008213_161 16100001 16210000
AP008213_162 16200001 16310000
AP008213_163 16300001 16410000
AP008213_164 16400001 16510000
AP008213_165 16500001 16610000
AP008213_166 16600001 16710000
AP008213_167 16700001 16810000
AP008213_168 16800001 16910000
AP008213_169 16900001 17010000
AP008213_170 17000001 17110000
AP008213_171 17100001 17210000
AP008213_172 17200001 17310000
AP008213_173 17300001 17410000
AP008213_174 17400001 17510000
AP008213_175 17500001 17610000
AP008213_176 17600001 17710000
AP008213_177 17700001 17810000
AP008213_178 17800001 17910000
AP008213_179 17900001 18010000
AP008213_180 18000001 18110000
AP008213_181 18100001 18210000
AP008213_182 18200001 18310000
AP008213_183 18300001 18410000
AP008213_184 18400001 18510000
AP008213_185 18500001 18610000
AP008213_186 18600001 18710000
AP008213_187 18700001 18810000
AP008213_188 18800001 18910000
AP008213_189 18900001 19010000
AP008213_190 19000001 19110000
AP008213_191 19100001 19210000
AP008213_192 19200001 19310000
AP008213_193 19300001 19410000
AP008213_194 19400001 19510000
AP008213_195 19500001 19610000
AP008213_196 19600001 19710000

```

```

Query Match 49.5%; Score 30.2; DB 15; Length 110000;
Best Local Similarity 74.5%; Pred. No. 1.1e+02;
Matches 38; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

```

```

11 GAAGAACTGCTCAAGAAACGCTGCTCAAGAAAGCTGCTCAAGAACTGTA 61

```

```

Db 71010 GATGAAGCTGTGAAGAGCAGCTGCTGTGAAGCTATGAGAACTGGA 70960
RESULT 10
AP005779/c 126938 bp DNA linear PLN 22-JUL-2004
LOCUS Oryza sativa [japonica cultivar-group] genomic DNA, chromosome 7,
DEFINITION BAC clone:OSJNB0042J07.
ACCESSION AP005779.3 GI:50508900
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Sasaki,T., Matsumoto,T. and Katayose,Y.
TITLE Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, BAC
clone:OSJNB0042J07
JOURNAL Published Only in Database (2002)
REFERENCE
AUTHORS Sasaki,T., Matsumoto,T. and Katayose,Y.
TITLE Direct Submission
JOURNAL Submitted (25-SEP-2002) Takuji Sasaki, National Institute of
Agrobiological Sciences, Rice Genome Research Program, Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaek@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
Tel:81-298-38-7441, Fax:81-298-38-7468)
On Jul 22, 2004 this sequence version replaced gi:33980494.
Genes were predicted from the integrated results of the following:
GENSCAN (http://CCR-081.mit.edu/GENSCAN.html), FGENESH
(http://www.softberry.com/), Genemark.hmm
(http://opal.biology.gatech.edu/Genemark/), glimmerM
(http://www.tigr.org/tdb/glimmerM/glmr_form.html), RiceHMM
(http://rgp.dna.affrc.go.jp/RiceHMM/), SplicePredictor
(http://bioinformatics.iaastate.edu/cgi-bin/sp.cgi), sim4
(http://globin.cse.psu.edu/html/docs/sim4.html), gap2
(http://www.tigr.org/software/glimmerM/), BLASTN and BLASTX. The
genomic sequence was searched against NCBI NonRedundant Protein
database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA
sequence database at RGP or DBJ. Protein homologues of the coding
regions were searched against NCBI NonRedundant Protein database
with BLASTP. ESTs represent the identified cDNA sequences using
BLASTN with the corresponding DBJ accession no. and RGP clone ID.
Full-length cDNAs represent the identified cDNA sequences using
BLASTN with the corresponding DBJ accession no.
A gene with identity or significant homology to a protein is
classified based on the protein name to indicate the homology level
such as same name, 'putative-' and '-like protein'. A gene without
significant homology to any protein but with full-length cDNA or
EST homology (covering almost the entire length of partial
sequence) is classified as an 'unknown' protein. A gene predicted
by two or more gene prediction programs is classified as a
'hypothetical' protein according to IRSP standard. A gene
predicted by a single gene prediction program is also classified as
a probable 'hypothetical' protein and is included as a
miscellaneous feature of the sequence.
The orientation of the sequence is from M13rev to -21M13 of the BAC
clone. This sequence of OSJNB0042J07 clone has an overlap with
OSJNB00008C11 (DBJ: AP005098) clone at 5' end and with OJ1664 D08
(DBJ: AP003931) at 3' end. Detailed information on overlap and
assembly quality together with annotation of this entry is
available at http://rgp.dna.affrc.go.jp/GenomesSeq.html.
FEATURES
source
Location/Qualifiers
1..126938
/organism="Oryza sativa [japonica cultivar-group]"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="7"
/clone="OSJNB0042J07"

```

gene complement(join(7428..7613,7693..7791,7866..7958,8043..8153,8250..8423,8533..9632,9707..11128))
/gene="OSJNB0042J07.1"
/complement(join(7428..7613,7693..7791,7866..7958,8043..8153,8250..8423,8533..9632,9707..11128))
/gene="OSJNB0042J07.1"
/note="TNP2-like transposable element-like"
complement(12045..12392)
/gene="OSJNB0042J07.2"
complement(12045..12392)
/gene="OSJNB0042J07.2"
/note="start and end point are not identified"
complement(12045..12392)
/gene="OSJNB0042J07.2"
/note="similar to Oryza sativa chromosome 10, OSJNB004E08.3"
/codon_start=1
/product="hypothetical protein"
/protein_id="BAC84457.1"
/db_xref="GI:34394149"
/translation="MPPPRHAAATPPHATATPPRRRHGPTTPRRRANRCDDAT
PPLTNNENDRMNSENHVVVRCRDVNNENNDRKTNSENDRKTSSENGTIVNE
LGNVETERNEDER"
12569..12772
/gene="OSJNB0042J07.3"
12569..12772
/gene="OSJNB0042J07.3"
/note="hypothetical ORF
predicted by Glimmerm
this category is not included in IRGSP standard"
complement(13986..16181)
/gene="OSJNB0042J07.4"
complement(13986..16181)
/gene="OSJNB0042J07.4"
/note="Rim2 protein"
complement(17731..17895)
/gene="OSJNB0042J07.5"
complement(17731..17895)
/gene="OSJNB0042J07.5"
/note="hypothetical ORF
predicted by Glimmerm
this category is not included in IRGSP standard"
complement(22300..23339,2475..23553))
/gene="OSJNB0042J07.6"
complement(join(22300..23339,2475..23553))
/gene="OSJNB0042J07.6"
/note="hypothetical ORF
predicted by GENSCAN
this category is not included in IRGSP standard"
25154..26197
/gene="OSJNB0042J07.7"
25154..26197
/gene="OSJNB0042J07.7"
/note="Bcat and end point are not identified"
25154..26197
/gene="OSJNB0042J07.7"
/codon_start=1
/product="putative cyclokinin synthase"
/protein_id="BAC84458.1"
/db_xref="GI:34394150"
/translation="MATSLSLAKPKPAVAVAALPRLVPPSIDMSALSPPELVSV
SRSMVAKRAVVMGATGTRKRLAVDLALQGEVITNDKQLHRLGDLVANAKVAD
EBAAGVHHLIGVAHPDEEFTADFRASRAAAVARGALPIIAGSSSYEEVDG
DBRAPRDYDCCETWVDLPVHGFVGRVDMCGRGVAIEIAFDPRDYSYG
WRAIGPELDATYLRSCAAGGEERRALANAEIDIKATPRMISGQRKIVRLRW
RIRRVATATFRRRGAAANEAMERHVAAPSIDVNSFLHGEFTTAETTAAPPPPP
LRFVFAAAGAGV"
complement(join(29815..29905,30289..30325,30436..30466,31631..31663))
/gene="OSJNB0042J07.8"
complement(join(29815..29905,30289..30325,30436..30466,31631..31663))
/gene="OSJNB0042J07.8"

/note="hypothetical ORF
predicted by GENSCAN
this category is not included in IRGSP standard"
join(32410..32415,33280..33507)
/gene="OSJNB0042J07.9"
join(32410..32415,33280..33507)
/gene="OSJNB0042J07.9"
/note="hypothetical ORF
predicted by GENSCAN
this category is not included in IRGSP standard"
complement(join(43221..43669,43754..43818,44701..44809,45640..45785,45988..46053,47141..47233,47756..47861,48542..48618,50293..50644))
/gene="OSJNB0042J07.10-1"
complement(join(43221..43669,43754..43818,44701..44809,45640..45785,45988..46053,47141..47233,47756..47861,48542..48618,50293..50644))
/gene="OSJNB0042J07.10-1"
/note="supported by full-length cDNA(s) : AK065959"
complement(43300..49894)
/gene="OSJNB0042J07.10-2"
complement(join(43300..43669,43754..43818,44701..44809,45640..45785,45988..46053,47141..47233,47756..47861,48542..48618,50293..50644))
/gene="OSJNB0042J07.10-2"
/note="supported by full-length cDNA(s) : AK065682"
complement(43300..49894)
/gene="OSJNB0042J07.10-2"
/note="contains full-length cDNA(s) : AK065682
non-coding transcript
probably inactive due to including stop codon(s) in CDS"
complement(join(43522..43669,43754..43818,44701..44809,45640..45785,45988..46053,47141..47233,47756..47861,48542..48618,50293..50653))
/gene="OSJNB0042J07.10-1"
/note="contains EST(s) : C72879(B2411)
contains full-length cDNA(s) : AK065959
similar to Arabidopsis thaliana chromosome 5, At5g19050.1"
/codon_start=1
/product="unknown protein"
/protein_id="BAD31697.1"
/db_xref="GI:50508901"
/translation="MTSSPSWFSGIARASSAMPPEGVAAAPLSDGAGRGCGGGG
GGGGVAAVAAAGAGAGAGAGGRRQVQALPKYKPAKQVAFKFGDFHGYIFI
GGLTDPLATYLPRLSLALYERKSLVQPLSSYTYGSLSEDLLELDLISL
INKNSDVIILGHSFGQDVLVHRTKIPACAKAAGVILQAPVSDREKATLPETAE
MIDLAAXMLSRGMDLMPREANPDAPITAYRHLCAVMGDDMFSSDLSDLROR
LGHSTTCOVYFSWGDVEYEVYKALVRLCRALGNAEKVEIEMGNHALSNVOE
AVRAIVDFVKKRGPKGMDPMS"
complement(53267..56811)
/gene="OSJNB0042J07.11"
complement(join(53267..54056,54177..54400,54826..55001,55788..56811))
/gene="OSJNB0042J07.11"
/note="start and end point are not identified"
complement(join(53267..54056,54177..54400,54826..55001,55788..56811))
/gene="OSJNB0042J07.11"
/codon_start=1
/product="putative bZIP protein"

Query Match 49.5%; Score 30.2; DB 15; Length 126938;
Best Local Similarity 74.5%; Pred. No. 1,1e+02;
Matches 38; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
CY 11 GAAGAACTGCTCAAGAACCGCTGCTCAAGAGCTGCTCAAGAACTGTA 61
DB 74020 GATGAACCTGTGAGAGGCACTGCTGCTGAAGCTATTGAAGAACTGGA 73970
RESULT 11
AP003931/c
LOCUS AP003931 128553 bp DNA linear PLN 22-JUL-2004

DEFINITION	Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 7, BAC clone:OJ1664_D08.
ACCESSION	AP003931
VERSION	AP003931.3
KEYWORDS	GI:50508415
SOURCE	Oryza sativa (japonica cultivar-group)
ORGANISM	Oryza sativa (japonica cultivar-group)
REFERENCE	1. Sasaki, T., Matsumoto, T. and Yamamoto, K. Oryza sativa nipponbare(GM3) genomic DNA, chromosome 7, BAC clone:OJ1664_D08 2. Sasaki, T., Matsumoto, T. and Yamamoto, K. Direct Submission Submitted (18-JUL-2001) Takuji Sasaki, National Institute of Agricultural Sciences, Rice Genome Research Program, Kamondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rsgp.dna.affrc.go.jp/, Tel:81-298-38-7441, Fax:81-298-38-7468) On JUL 22, 2004 this sequence version replaced gi:22531714. Genes were predicted from the integrated results of the following: GENSCAN (http://CCR-081.mit.edu/GENSCAN.html), FGENESH (http://www.softberry.com/), Genemark (http://opal.biology.gatech.edu/Genemark/), Glimmer (http://www.tigr.org/tdb/glimmer/glimmerform.html), RiceHMM (http://rsgp.dna.affrc.go.jp/RiceHMM/), SplicePredictor (http://bioinformatics.isabate.edu/cgi-bin/sp.cgi), SIm4 (http://globin.cse.psu.edu/html/docs/sim4.html), gap2 (http://www.tigr.org/software/glimmer/), BLASTN and BLASTX. The genomic sequence was searched against NCBI NonRedundant Protein database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at RGP or DBJ. Protein homologs of the coding regions were searched against NCBI NonRedundant Protein database with BLASTP. ESTs represent the identified cDNA sequences using BLASTN with the corresponding DBJ accession no. and RGP clone ID. Full-length cDNAs represent the identified cDNA sequences using BLASTN with the corresponding DBJ accession no.
COMMENT	A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level such as same name, 'putative-' and '-like protein'. A gene without significant homology to any protein but with full-length cDNA or EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted by two or more gene prediction programs is classified as a 'hypothetical' protein according to IRGSP standard. A gene predicted by a single gene prediction program is also classified as a probable 'hypothetical' protein and is included as a miscellaneous feature of the sequence. The orientation of the sequence is from -21M13 to M13rev of the BAC clone. This sequence of OJ1664_D08 clone has an overlap with OSUNDB0042J07 (DBJ: AP005779) at 5' end and with OJ116_C08 (DBJ: AP004002) at 3' end. The sequence was generated by combining Monsanto and RGP-Japan sequencing data. Detailed information on overlap and assembly quality together with annotation of this entry is available at http://rsgp.dna.affrc.go.jp/GenomeSeq.html.
FEATURES	Source 1..128553 /organism="Oryza sativa (japonica cultivar-group)" /mol_type="genomic DNA" /cultivar="Nipponbare" /db_xref="taxon:39947" /chromosome="7" /clone="OJ1664_D08" complement(join(1360..1450,1834..1870,1981..2011,3176..3208)) /gene="OJ1664_D08_101" complement(join(1360..1450,1834..1870,1981..2011,3176..3208)) /gene="OJ1664_D08_101"
misc_feature	misc_feature /note="hypothetical ORF predicted by GENSCAN this category is not included in IRGSP standard" join(3955..3960,4825..5052) /gene="OJ1664_D08_102" join(3955..3960,4825..5052) /gene="OJ1664_D08_102" /note="hypothetical ORF predicted by GENSCAN this category is not included in IRGSP standard" complement(join(14766..15214,15299..15363,16246..16354,17185..17330,17533..17598,18686..18778,19301..19406,20087..20163,21838..22189)) /gene="OJ1664_D08_103-1" complement(join(14766..15214,15299..15363,16246..16354,17185..17330,17533..17598,18686..18778,19301..19406,20087..20163,21838..22189)) /gene="OJ1664_D08_103-2" complement(join(14845..21439)) /gene="OJ1664_D08_103-3" complement(join(14845..21439)) /note="supported by full-length cDNA (s) : AK065682" complement(14845..21439) /gene="OJ1664_D08_103-4" complement(join(115067..15214,15299..15363,16246..16354,17185..17330,17533..17598,18686..18778,19301..19406,20087..20163,21838..22098)) /gene="OJ1664_D08_103-1" /note="contains EST(s) : C72879(E2411) contains full-length cDNA(s) : AK065559 similar to Arabidopsis thaliana chromosome 5, At5g19050.1" /codon_start=1 /product="unknown protein" /protein_id="BAC83224.1" /db_xref="GI:34393295" /translation="MTSSPSWFGIRARASSAMPFGVSAAPAPLSDAGRGCGGGG/GGGGVAAVAAPGAGVGGGRKQQLFYGKSAQAVAKTQDPHQVFI/GGLDFGLADYDLSLLEVKMSLVQPLSSVTGSISSLEQDALELDLISL/INKENSDGVLLIHSTGCODIYVMYKTRACSAVGVLIQAPVSRERATLPEAE/MIDLAAMLSSEGMDMPREAPDAPIYAYRHSICAVMGDDMFSSDLSBQLROR/LGHMSTCCOVITPSMGDEYVPEVVDKALVDRCRLGNAEKYEIMGNHALSNRQE/AVAIYDFVRSEPKMGDDPWS" complement(24812..28356) /gene="OJ1664_D08_104" complement(24812..25601,25722..25945,26371..26546,27333..328356)) /gene="OJ1664_D08_104" /note="start and end point are not identified" complement(join(24812..25601,25722..25945,26371..26546,27333..28356)) /gene="OJ1664_D08_104" /codon_start=1 /product="putative bZIP protein" /protein_id="BAC16409.1" /db_xref="GI:23237834" /translation="MGCAOSNEDGEGPVARCERKKILRDVAVARHALGATAGAA/LKVGAALSDYASGEGEALRGSGADSSAAALATASDGKPRVLAIPRPP/LPPPPRPLPRHGDVAAPLARSMSRDLNHLQIKKRGSGAPRIEEDDEGCGG/DAGRGDDDAELKPRPLPSORPPRSPPRLPRPDHKKVDRPGGFISSLPSMPR/TLDTAALEPSSSAERREPPRPAPRDEHDSAAAEVAVAGSKRAAABATTPRAAMOK/AARKGAKAAMLVAAPPOPAKLVGDITLALDENFLKASQSAHEVSKLEAKRMVH/SNFAETRGFDHSARVQVITWNRSFKGIQPEPVNENDDMEHTATLDTLAME/KKLHNEKVDPEVLRREYQGLKANKKGGVSSLEKTSVSHLHTKYVVDLOTM/ESTVAETNRLRDOOLYKLELTVKGMKMNDAVLAHNKQIKLILEKSLDISVARE/TSEGNHRTYQLMNVNENHTQDKEMTYQKQVSGIYPTIKLVNIPDNILKPSNSO/PHEITPRIKVLAMNEITGLRPDEAKAKAINFPAITVKILVHQBDELKRLKIKIEDT

	gene	EMVYGGKRRQFDPMACKYHINOTGAILPEDNNPOCARPDPMARCAANEKELSKMELE BRWYKQCRVRYREKSLSILRTNIPDLFRVAVDFSLQSAGMGFGWSIAHTNDQULE"
	repeat_unit	/gene="OU1664_D08.105" 30227. .36900 30227. .30585 /gene="OU1664_D08.105" /note="3' terminal repeat" /rpt_type=terminal 31648. .36161 /gene="OU1664_D08.105" /note="probably inactive due to including stop codon(s) in CDS pseudo gene, copia polyprotein"
	misc_feature	36539. .36900 /gene="OU1664_D08.105" /note="3' terminal repeat" /rpt_type=terminal complement(join(39517. .39634,39757. .39812,39934. .39987, 40221. .40289)) /gene="OU1664_D08.106" complement(join(39517. .39634,39757. .39812,39934. .39987, 40221. .40289)) /gene="OU1664_D08.106" /note="hypothetical ORF predicted by GeneMark.hmm this category is not included in IRGSP standard"
	gene	40722. .43447 /gene="OU1664_D08.107" join(40722. .41153,41226. .41350,41434. .41536,41644. .41782 42196. .44273,43003. .43447) /gene="OU1664_D08.107" /note="supported by full-length cDNA(s)" join(40723. .41153,41226. .41350,41434. .41536,41644. .41782 42196. .44273,43003. .43438) /gene="OU1664_D08.107" /note="supported by full-length cDNA(s)" join(40785. .41153,41226. .41350,41434. .41536,41644. .41782 42196. .44273,43003. .43167) /gene="OU1664_D08.107" /note="contains EST(s)": D48305(S14442), AU096613(S14442) contains full-length cDNA(s)": AK071575,AK104057" /codon_start=1 /product="putative mRNA binding protein precursor" /protein_id="BAC83225.1" /db_xref="GI:34393296"
	mRNA	/translation="MAFSPTAAASGATFAVSGGHRRGACHLLSSFLIAPPEVAARR CGEAAAARRRGAATVRQAQAAGAAGKESVLIVNTMGCHAVIGFYLAIDLAAAGVAT VLTVGDESDMKRPFSRSESLTSAGATTWGDDPADVGAAVGGASDVTIDNRKD LDAYKPVVDMAKAAGVNAQFLFVSAGITTPSDEPRPHVGDADAKESAGHVGEKYIAEQ FGSAVSPQVMIGSGNQDKCEMFEDRIVRGRPVPVIGSGMQVNTNISHRVRLASMAVA LAVESPGAAGRIFPCVCDRAVTFNGLVLMCAAAGAOBEILHYDPAAVGVDAKRAF FRNHVFAPRAAKEVLDRSRSTNPDLCKEFAPYAASGRQKEMSFDLDKIITAAA
	CDS	
Query Match	49.5%; Score 30.2; DB 15; Length 128553;	
Best Local Similarity	74.5%; Pred. No. 1.e+02;	
Matches	38; Conservative 0; Mismatches 13; Indels 0; Gaps 0;	
Db	11 GAAGAACTGCTCAGAAACCGCTCTCAAGAGCTGCACAGAAACTGTGA 61 45565 GATGAAACTGTGAAGAGCGCACTGCTCTGTAAGCTATTGAAGAACTGGA 45515	
RESULT 12		
PFAILERC		
LOCUS	PFAILERC 414 bp DNA linear INV 26-APR-1993	
DEFINITION	P.chabaudi merozoite surface antigen 1 (MSA-1) gene, partial cds.	
ACCESSION	M61208	
VERSION	M61208.1 GI:160022	
KEYWORDS	merozoite surface antigen 1.	
SOURCE	Plasmodium chabaudi	
ORGANISM	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.	

	REFERENCE	1 (bases 1 to 414)
	AUTHORS	Deleersnijder W., Hendrix D. and Hamers R.
	TITLE	Analysis of MSA-1 diversity in Plasmodium chabaudi strains
	JOURNAL	Mol. Biochem. Parasitol. 46 (2), 315-317 (1991)
	PUBMED	1922203
	COMMENT	Original source text: Plasmodium chabaudi (strain 11R) DNA.
	FEATURES	location/Qualifiers
	source	1..414
	gene	/organism="Plasmodium chabaudi"
		/mol_type="genomic DNA"
		/strain="11R"
		/db_xref="taxon:5825"
	CDS	1..414
		/gene="MSA-1"
		1..414
		/gene="MSA-1"
		/codon_start=3
		/product="merozoite surface antigen 1"
		/protein_id="AA29453.1"
		/db_xref="GI:552165"
		/translation="APRPVAPETAPETAPEPAEPAPQEPQESAPAEATET
		TPPESASTPEPPAPPTPSSETVTGEGTTPAPKPKQEGASSAPAPAPAKPAPQ
		TVTGGSTNVEGSTOVRASEDEMFVDDEFVNPFYK"
ORIGIN		
Query Match	48.9%; Score 29.8; DB 2; Length 414;	
Best Local Similarity	75.5%; Pred. No. 2.2e+02; Mismatches 12; Indels 0; Gaps 0;	
Matches	37; Conservative	
Oy	13 AGAACTGCTCAAGAACCGTGCTCAAGAAGTGTCTCAAGAACTGTA 61 53 AGAACTGCACCAAAACACCAGCACAAGAGCTCCACAACACAGA 101	
Dd	53	
RESULT 13		
PFASAC		
LOCUS	414 bp DNA linear INV 26-APR-1993	
DEFINITION	P.chabaudi merozoite surface antigen 1 (MSA-1) gene, partial cds.	
ACCESSION	M61206 GI:160099	
VERSION	M61206.1	
KEYWORDS	merozoite surface antigen 1.	
SOURCE	Plasmodium chabaudi	
ORGANISM	Plasmodium chabaudi	
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.	
AUTHORS	1 (baes 1 to 414)	
TITLE	Deleersnijder W., Hendrix D. and Hamers R.	
JOURNAL	Analysis of MSA-1 diversity in Plasmodium chabaudi strains	
PUBMED	Mol. Biochem. Parasitol. 46 (2), 315-317 (1991)	
COMMENT	1922203	
FEATURES	Original source text: P.chabaudi (strain AS) DNA. location/Qualifiers	
source	1..414	
	/organism="Plasmodium chabaudi"	
	/mol_type="genomic DNA"	
	/strain="AS"	
	/sub_strain="chabaudi"	
	/db_xref="taxon:5825"	
	1..414	
	/gene="MSA-1"	
	<1..>414	
	/gene="MSA-1"	
	/codon_start=3	
	/product="merozoite surface antigen 1"	
	/protein_id="AAA29494.1"	
	/db_xref="GI:552183"	
	/translation="APRPVAPETAPETAPEPAEPAPQEPQESAPAEATET	
	TPPESASTPEPPAPPTPSSETVTGEGTTPAPKPKQEGASSAPAPAPAKPAPQ	
	TVTGGSTNVEGSTOVRASEDEMFVDDEFVNPFYK"	
ORIGIN		
Query Match	48.9%; Score 29.8; DB 2; Length 414;	
Best Local Similarity	75.5%; Pred. No. 2.2e+02; Mismatches 12; Indels 0; Gaps 0;	
Matches	37; Conservative	

